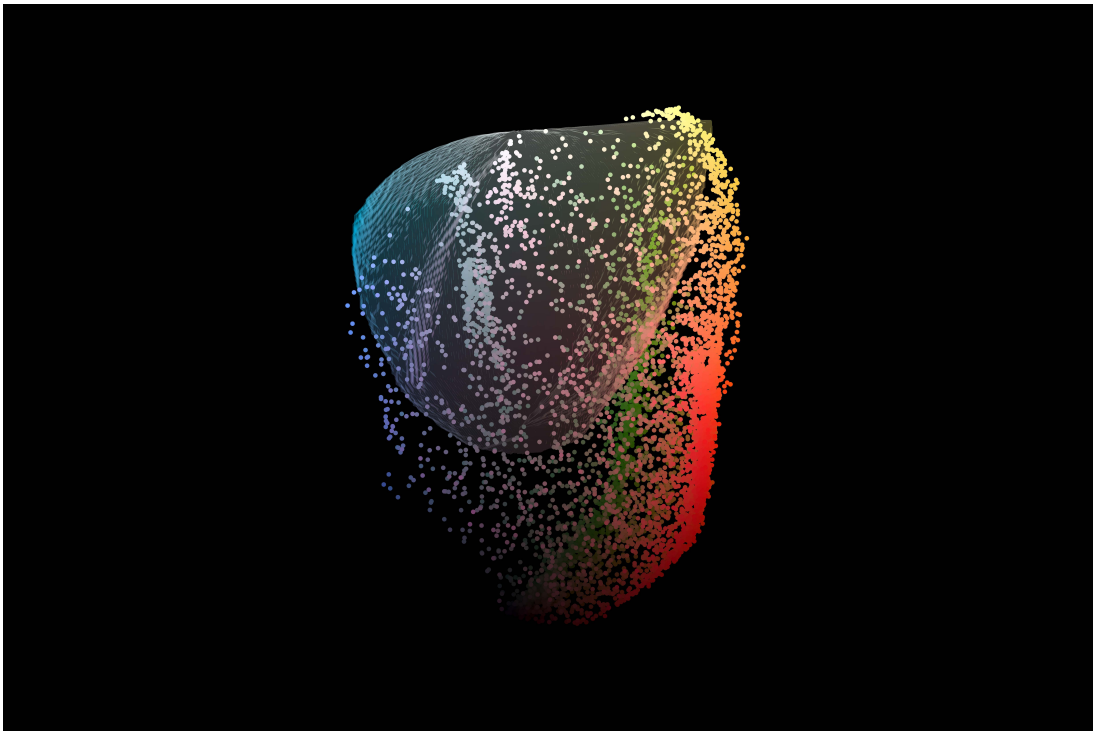
Color-Image Quality Assessment: From Metric to Application



Jens Preiss

Color-Image Quality Assessment: From Metric to Application

Vom Fachbereich Maschinenbau
an der Technischen Universität Darmstadt
zur
Erlangung des akademischen Grades eines
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte

Dissertation

vorgelegt von
Dipl.-Phys. Jens Preiss
aus Reutlingen

Berichterstatter: Dr. rer. nat. Philipp Urban
Prof. Dr.-Ing. Edgar Dörsam
Prof. Dr.-Ing. Michael Goesele

Datum der Einreichung: 18.08.2014
Datum der mündlichen Prüfung: 17.12.2014

Darmstadt 2015
D17

Kurzfassung

Die Bewertung der visuellen Qualität von Bildern in der digitalen Bildverarbeitung ist eine entscheidende Voraussetzung für die meisten bildverarbeitenden Systeme. Für eine solche Bewertung der Bildqualität werden hauptsächlich objektive Bewertungen, welche automatisch die Bildqualität durch einen Computeralgorithmus voraussagen, verwendet. Die große Mehrheit der objektiven Bewertungen sind sogenannte *Bildabstandsmetriken*. Diese sagen den wahrgenommenen Unterschied zwischen einem verzerrtem Bild und einer Referenz voraus. Aufgrund des beschränkten Verständnisses des menschlichen visuellen Systems ist die Bewertung der Bildqualität kompliziert und noch immer ein offenes Forschungsfeld.

Die Mehrheit der Bildabstandsmetriken vernachlässigt Farbinformationen, was schnellere Berechnungen zulässt. Auch wenn ihre Leistungsfähigkeit für viele Anwendungen ausreichend ist, sind die Metriken nicht in der Lage, die Qualität für eine Vielfalt an Farbverzerrungen korrekt vorauszusagen. Viele Bildabstandsmetriken berücksichtigen außerdem nicht die Betrachtungsbedingungen, welche einen großen Einfluss auf die wahrgenommene Bildqualität haben können (z. B. ein großer Bildschirm in einem Büro verglichen mit einem kleinen Mobilgerät im hellen Sonnensschein).

Das Hauptziel meiner Forschung war die Entwicklung einer neuen Bildabstandsmetrik, genannt *iCID*, die Bilder auf Standardbetrachtungsbedingungen normalisiert und chromatische Merkmale extrahiert. Die neue Metrik wurde dann als Zielfunktion eingesetzt, um sowohl *Gamut-Mapping* als auch *Tone-Mapping* zu verbessern. Beide Methoden stellen wesentliche Transformationen zur Farbbildwiedergabe dar.

Die Leistungsfähigkeit der vorgeschlagenen Metrik wurde durch visuelle Experimente als auch durch Vergleiche mit subjektiven Beurteilungen bestätigt. Die visuellen Experimente offenbaren signifikante Verbesserungen gegenüber hochmodernen Gamut-Mapping- und Tone-Mapping-Transformationen. iCID weist für Gamut-Mapping-Verzerrungen die signifikant höchste Korrelation zu subjektiven Beurteilungen auf und übertrifft für konventionelle Verzerrungen (z. B. Rauschen, Unschärfe und Kompressionsartefakte) fast alle hochmodernen Metriken.

Abstract

In digital imaging, evaluating the visual quality of images is a crucial requirement for most image-processing systems. For such an image quality assessment, mainly objective assessments are employed which automatically predict image quality by a computer algorithm. The vast majority of objective assessments are so-called *image-difference metrics* which predict the perceived difference between a distorted image and a reference. Due to the limited understanding of the human visual system, image quality assessment is not straightforward and still an open research field.

The majority of image-difference metrics disregard color information which allows for faster computation. Even though their performance is sufficient for many applications, they are not able to correctly predict the quality for a variety of color distortions. Furthermore, many image-difference metrics do not account for viewing conditions which may have a large impact on the perceived image quality (e.g., a large display in an office compared with a small mobile device in the bright sunlight).

The main goal of my research was the development of a new image-difference metric called *improved Color-Image-Difference* (iCID) which normalizes images to standard viewing conditions and extracts chromatic features. The new metric was then used as objective function to improve *gamut mapping* as well as *tone mapping*. Both methods represent essential transformations for the reproduction of color images.

The performance of the proposed metric was verified by visual experiments as well as by comparisons with human judgments. The visual experiments reveal significant improvements over state-of-the-art gamut-mapping and tone-mapping transformations. For gamut-mapping distortions, iCID exhibits the significantly highest correlation to human judgments and for conventional distortions (e.g., noise, blur, and compression artifacts), iCID outperforms almost all state-of-the-art metrics.

Acknowledgments

First of all, I would like to thank my supervisor Dr. Philipp Urban who advised me during my graduation and taught me how to be a good scientist. In close collaboration with him, I had the best conditions to successfully pass the time of my graduation. Prof. Edgar Dörsam as the head of our institute provided an extraordinarily productive and positive environment which allowed me to fully focus on my research. I also thank Prof. Michael Goesele who did not hesitate to review this thesis for spending so much of his valuable time.

This thesis would not have been possible without my wife Yvonne who helped me immensely with her infinite patience, understanding and sacrifice. It is invaluable to have someone you can rely on when you need it the most. I am deeply grateful to her and my family for their continuous support at any given moment.

Finally, I wish to thank my colleagues at the Institute of Printing Science and Technology for having a great time inside and outside the institute. They have always been helpful, sharing their knowledge, and — good friends. I am especially indebted to the members of the color group for the successful collaboration.

This work was supported by the German Research Foundation within the Emmy-Noether Program of "Perceptually Optimal Reproduction of Color Images considering Device Limits".

Darmstadt, February 2015

Contents

Notations

Acronyms and Abbreviations

CID	**C**olor **I**mage **D**ifference; name of an image-difference metric
CIE	**C**ommission **I**nternationale de l'**É**clairage; International Commission on Illumination
CIECAM02	**C**olor-**A**ppearance **M**odel proposed by the CIE in 2002
CIEDE2000	color-difference formula on the CIELAB color space
CIELAB	CIE 1976 color space with color-opponent coordinates (L^*, a^*, b^*)
CIEXYZ	CIE color space with tristimulus values $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$
CRT	**C**athode **R**ay **T**ube; display technology (out of date)
CSF	**C**ontrast-**S**ensitivity **F**unction
Empa	**E**idgenössische **M**aterial**P**rüfungs- und Forschungs-**A**nstalt; Swiss Federal Laboratories for Materials Science and Technology
FSIM	**F**eature **SIM**ilarity; name of an image-difference metric
GMA	**G**amut-**M**apping **A**lgorithm
HDR	**H**igh **D**ynamic **R**ange; luminance range closer to real-world scenes
hdr-LAB2000HL	**h**igh-**d**ynamic-**r**ange extension of the LAB2000HL color space
HVS	**H**uman **V**isual **S**ystem
IAM	**I**mage-**A**ppearance **M**odel; image model accounting for viewing conditions
ICC	**I**nternational **C**olor **C**onsortium; develops international standards for color management
iCID	**i**mproved **C**olor **I**mage **D**ifference; name of an image-difference metric and improved version of CID
IDF	**I**mage-**D**ifference **F**eature
IDM	**I**mage-**D**ifference **M**etric
IQA	**I**mage **Q**uality **A**ssessment
JND	**J**ust-**N**oticeable **D**ifference; threshold of perception

JPEG	**J**oint **P**hotographic **E**xperts **G**roup; image file format which uses lossy compression for storage
LAB2000HL	hue linear and nearly perceptually uniform color space
LCD	**L**iquid-**C**rystal **D**isplay; common display technology
LDR	**L**ow **D**ynamic **R**ange; luminance range of most devices
LIVE	**L**aboratory for **I**mage and **V**ideo **E**ngineering; LIVE is located at The University of Texas at Austin
MATLAB	**MAT**rix-**LAB**oratory; a numerical computing environment by MathWorks
MOS	**M**ean **O**pinion **S**cores; subjective scores obtained by a visual experiment
MSE	**M**ean **S**quared **E**rror; name of an image-difference metric
MSSIM	**M**ultiscale **S**tructural **SIM**ilarity; multiscale version of the SSIM index
NTSC	**N**ational **T**elevision **S**ystem **C**ommittee; analog television system
PSNR	**P**eak **S**ignal-to-**N**oise **R**atio; name of an image-difference metric
RGB	**R**ed-**G**reen-**B**lue; additive color model based on the three primaries red, green, and blue
sRGB	**s**tandard **R**ed-**G**reen-**B**lue; standardized RGB color space
SSIM	**S**tructural **SIM**ilarity; name of an image-difference metric
TID	**T**ampere **I**mage **D**atabase; image-quality-assessment database comprising conventional distortions
TMO	**T**one-**M**apping **O**perator
UQI	**U**niversal Image **Q**uality **I**ndex; name of an image-difference metric
VIF	**V**isual **I**nformation **F**idelity; name of an image-difference metric

Symbols

a	red-green channel in LAB2000HL color space
a^*	red-green channel in CIELAB color space
a_{HDR}	red-green channel in hdr-LAB color space
α	exponents to weight the CID, iCID, and SSIM components
b	blue-yellow channel in LAB2000HL color space

b^*	blue-yellow channel in CIELAB color space
b_{HDR}	blue-yellow channel in hdr-LAB color space
$\beta_1, ...\beta_n$	scale weights of IDFs for n scales
$\mathbf{C}$	contrast-feature extraction
C	chroma channel in LAB2000HL color space
C^*	chroma channel in CIELAB color space
$\mathbf{C}_L$	lightness-contrast IDF
CI	confidence interval of an estimated parameter
CID	prediction of the CID metric
c	contrast; image-difference map of the SSIM index
c_C	chroma-contrast comparison
c_L	lightness-contrast comparison
$c_1, ..., c_7$	constants of the SSIM, CID, and iCID metric
γ	significance level of an estimated population parameter
$\mathbf{D}$	downsampling transformation
D	viewing distance to a visual target; $[D]$ = [Length] = m
$\mathbf{D}_j$	sRGB image ($\mathbf{D}_j \in \mathcal{I}_{M,N}$) with non-zero values only at image position j, $j \in \{1, ..., M\} \times \{1, ..., N\}$
DR	dynamic range of an image; e.g., $DR = 255$ for an 8-bit image
$\mathbf{d}$	image-difference prediction without accounting for the viewing conditions
δ	side length of the cubes in the quantized LAB2000HL color space
ΔE^*_{ab}	color-difference formula; Euclidean distance in the CIELAB color space
$\mathbf{F}$	actual feature extraction after normalization
$\mathcal{G}$	device (output) gamut
$\mathbf{G}$	gamut-mapping transformation
$\mathcal{H}$	set of all in-gamut pixel deviations with step size δ in the quantized LAB2000HL color space
H	hue channel in LAB2000HL color space
H^*	hue channel in CIELAB color space
$\mathbf{h}$	human-population-average image-difference perception
$\mathcal{I}_{M,N}$	set of all sRGB images with M rows and N columns
IDF	image-difference feature extraction
IDM	image-difference prediction
iCID	prediction of the iCID metric
$\mathcal{K}$	all pairs of corresponding sliding windows within the normalized images
k	side length of a rectangular sliding window
$\mathbf{L}$	difference-feature extraction, formerly luminance-feature extraction

L	lightness; achromatic channel in LAB2000HL color space
L^*	achromatic channel (lightness) in CIELAB color space
$\mathbf{L}_C$	chroma-difference IDF
$\mathbf{L}_H$	hue-difference IDF
$\mathbf{L}_L$	lightness-difference IDF
L_{HDR}	achromatic channel (lightness) in hdr-LAB color space
l	luminance; image-difference map of the SSIM index
l_C	chroma-difference comparison
l_H	hue-difference comparison
l_L	lightness-difference comparison
M	number of rows of an image
m	number of scales used in a multi-scale approach
$\mu_{\mathbf{x}}, \mu_{\mathbf{y}}$	Gaussian-weighted mean for pixel arrays $\mathbf{x}$,$\mathbf{y}$
$\mathbf{N}$	image-normalization transformation
N	number of columns of an image
$\hat{n}$	number of all predictions of an image-difference metric
$\mathcal{P}$	set of all parameter arrays describing the viewing conditions
P	number of horizontal pixels on a screen; $[P]$ = samples = s
p	success probability of a population parameter
$\hat{p}$	hit rate; ratio of correct predictions to all predictions
$\hat{p}_m$	majority hit rate; maximal achievable hit rate
$\hat{p}_r$	hit-rate ratio for valuating a hit rate
Ψ	radius of a confidence interval
$\hat{r}$	number of correct predictions of an image-difference metric
$\mathcal{S}$	all pairs of corresponding sliding windows within a subimage of the images
$\mathbf{S}$	structure-feature extraction
$\mathbf{S}_L$	lightness-structure IDF
SF	spatial frequency of a visual target; $[SF]$ = cycles / degree of visual angle = c/deg
s	structure; image-difference map of the SSIM index
s_C	chroma-structure comparison
s_L	lightness-structure comparison
σ_G	standard deviation of the weighting function w
$\sigma_{\mathbf{x}}, \sigma_{\mathbf{y}}$	Gaussian-weighted standard deviations for pixel arrays $\mathbf{x}$,$\mathbf{y}$
$\sigma_{\mathbf{xy}}$	Gaussian-weighted covariance of pixel arrays $\mathbf{x}$ and $\mathbf{y}$
$\mathcal{T}$	all pairs of corresponding sliding windows within the total images
t	image-comparison transformation

VR	visual resolution of a display depending on the viewing distance and the pixel pitch; $[VR]$ = samples / degree of visual angle = s/deg
$\mathcal{W}$	all pairs of corresponding sliding windows within a $k \times k$-window
$\mathcal{W}_{M,N}$	set of all color images represented in a working color space with M rows and N columns
W	horizontal width of a screen; $[W]$ = [Length] = m
w	circular-symmetric Gaussian weighting with standard deviation σ_G
$\mathcal{X}$	original image of the gamut-mapping optimization
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$	sRGB images
X,Y,Z	tristimulus values of the CIEXYZ color space
X, Y	grayscale images
$\mathbf{X}_{Norm}, \mathbf{Y}_{Norm}$	images normalized to the viewing conditions represented in the working color space
$\mathbf{x}, \mathbf{y}$	pixel arrays of sRGB images
x, y	pixel arrays of grayscale images
$\mathcal{Y}$	starting image of the gamut-mapping optimization
Y	luminance; perceived brightness as well as linear grayscale values of an image
Y'	luma; perceived brightness as well as gamma-compressed grayscale values of an image
$\mathcal{Z}$	optimized image of the gamut-mapping optimization
$z_{\gamma/2}$	upper $\gamma/2$-th quantile of the standard normal distribution

Publications

As a research assistant at the Technische Universität Darmstadt from February 2011 to June 2014, I was involved in several journal and conference papers. The publications comprise:

Journal Papers

- J. Preiss, F. Fernandes, and P. Urban. Color-Image Quality Assessment: From Prediction to Optimization. *IEEE Transactions on Image Processing*, 23(3):1366–1378, 2014
- I. Lissner, J. Preiss, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. Image-Difference Prediction: From Grayscale to Color. *IEEE Transactions on Image Processing*, 22(2):435–446, 2013

Conference Papers

- S. Le Moan, J. Preiss, and P. Urban. Evaluating the multi-scale iCID metric. In *Electronic Imaging 2015: Image Quality and System Performance XII (accepted)*, 2015
- J. Preiss, M. D. Fairchild, J. A. Ferwerda, and P. Urban. Gamut Mapping in a High-Dynamic-Range Color Space. In *Proc. SPIE 9015, Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications*, p. 90150A, 2014
- J. Preiss and P. Urban. Image-Difference Measure Optimized Gamut Mapping. In *IS&T/SID, 20th Color and Imaging Conference*, pp. 230–235, 2012
- M. Scheller Lichtenauer, P. Zolliker, J. Preiss, and P. Urban. Image similarity for chromatic content. In *18. Workshop Farbbildverarbeitung*, pp. 107–117, 2012
- J. Preiss, I. Lissner, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. The Impact of Image-Difference Features on Perceived Image Differences. In *CGIV 2012 – 6th European Conference on Colour in Graphics, Imaging, and Vision*, pp. 43–48, 2012

PUBLICATIONS

- M. Scheller Lichtenauer, P. Zolliker, I. Lissner, J. Preiss, and P. Urban. Learning Image Similarity Measures from Choice Data. In *CGIV 2012 – 6th European Conference on Colour in Graphics, Imaging, and Vision*, pp. 24–30, 2012
- I. Lissner, J. Preiss, and P. Urban. Predicting Image Differences Based on Image-Difference Features. In *IS&T/SID, 19th Color and Imaging Conference*, pp. 23–28, 2011
- J. Preiss and P. Urban. Information Content Redundancies of Image Quality Assessments. In *17. Workshop Farbbildverarbeitung*, pp. 97–105, 2011

1

Introduction

Nowadays, digital color images are taken for granted in our daily life. We immediately post photos of what we are doing on the Internet and share them with friends, we go shopping online and see the product on the screen, we take hundreds of pictures on vacation with a digital camera and sort and process them afterwards.

In most of the cases, we do not have to take care of the quality of the images because image-processing technologies provide good quality even for the consumer market. Yet, we are sometimes disappointed by the results: the posted image looks impressive on the mobile device but has clearly visible artifacts when displayed on a large screen, such as noise, compression artifacts, and dull colors; the delivered product from an online shop does not have the desired colors as we saw them on our screen; the photo of the sunset at the beach is by no means as impressive as the sunset we witnessed during that moment.

Therefore, image quality is an essential factor in digital imaging, in general for color images. In each step of image processing, we seek the highest possible quality. However, what does quality mean and how do we assess quality of color images? This topic denoted as *color-image quality assessment* is treated in my thesis. Furthermore, novel contributions to color-image quality assessment are proposed which improve the current state-of-the-art and expand the field of possible applications.

1.1 Motivation

When images are altered through digital processing, assessing the impact of the alterations on the resulting visual quality is a crucial requirement for any image-processing system [19]. I.e., it is important to know in image processing how the *human visual system* (HVS) evaluates image quality. However, the HVS is far from being understood and each observer judges image quality differently. Therefore, the assessment of image quality is not straightforward and still an open research field.

I introduce the problem of image quality assessment for color images using a real-world example: a camera manufacturer has developed a new algorithm for denoising captured images but some parameters have to be adjusted for each camera model. To determine the model parameters, high-quality images without noticeable noise which serve as reference images are distorted by adding simulated noise of the camera model. Then, the aim is to find the parameter set which minimizes the perceived difference between the reference images and the images resulting from the denoising algorithm applied to the distorted images. Such a perceived image difference is referred to as relative-to-reference image quality assessment.

A quite exact way to assess image quality is to perform a visual experiment in which human observers are asked to judge image quality – a *subjective* assessment of image quality. Reliable results require a careful design of the experiment and its analysis, usage of high performance devices, and many observers. Such a procedure, however, is generally very time-consuming, cumbersome, and expensive. Hence, subjective assessment of image quality is usually not feasible for the camera manufacturer.

A much more efficient approach is to assess image quality automatically by a computer algorithm – an *objective* assessment of image quality. Such an algorithm which assesses perceived image differences is called *image-difference metric* (IDM). IDMs are usually fast, easy-to-use, and economic but are required to highly correlate with human judgments for satisfactory image-difference predictions.

Image processing is a subdiscipline of signal processing – an image can be considered as a discrete signal which contains the color information of all image pixels. In signal processing, the *mean-squared error* (MSE, for details see Section 3.3.1) is the "standard criterion for the assessment of signal quality and fidelity" [148]. Is the MSE then the right choice for assessing image quality in the mentioned problem? Most likely not, since the MSE shows "serious shortcomings, especially when dealing with perceptually important signals such as speech and images" [148]. In particular, image-difference predictions of the MSE hardly agree with the HVS as shown in Figure 1.1. Even though the MSE remains constant, the images exhibit clearly visible differences in quality.

Back to our camera example, Figure 1.1 can also be interpreted as a part of the mentioned denoising problem. Starting with the high-quality image (reference image), adding simulated noise to it yields the camera response (initial image). Then, varying the yet undetermined parameters of the denoising algorithm leads to the other images. The IDM can help to find the image with the minimal perceived difference to the reference image and thus to find the desired parameters.

As the MSE is not feasible for this task, a better performing IDM needs to be found. Research on IDMs has been done for decades resulting in more than 100 published IDMs [101]. A state-of-the-art IDM is the *Structural Similarity* (SSIM) index (for details see Section 3.3.2) which was used as the objective function in Figure 1.1. SSIM shows a high correlation to human judgments for a variety of distortion types such as noise, blur, or compression [129].

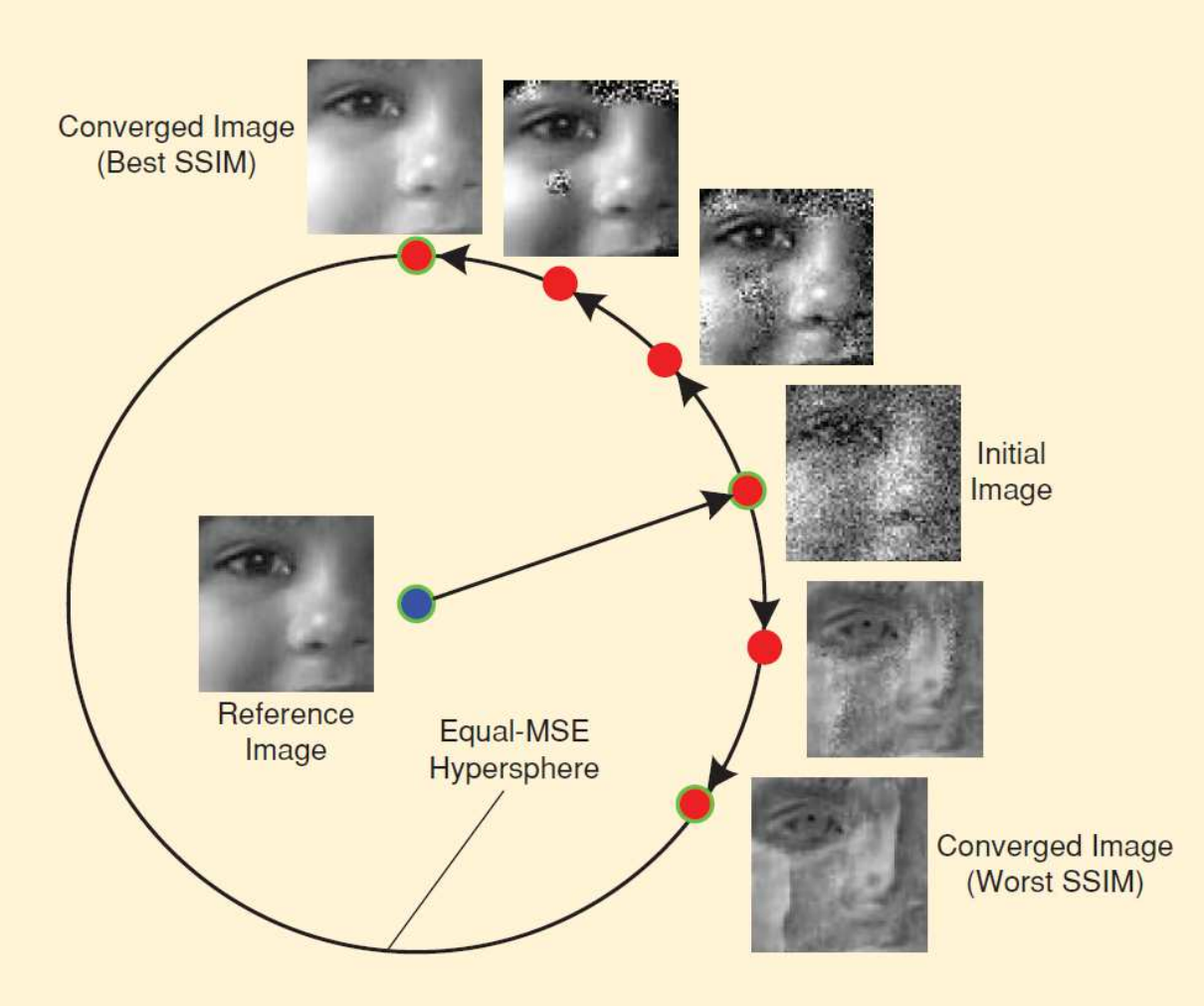


Figure 1.1: Finding the minimum and maximum of an objective function along the equal-MSE hypersphere. The reference image is first distorted yielding the initial image. Then, an iterative procedure minimizes and maximizes a fidelity metric (SSIM in particular) used as objective function keeping the MSE constant – in relation to the reference image. The figure has originally been published in [148].

The majority of IDMs – including SSIM – neglect color information for assessing image quality; they operate only on grayscale images (corresponding to intensity information). Such grayscale IDMs allow faster computation and may be sufficient for a variety of applications because grayscale images maintain structure of the images to a great extent. However, color plays an important role for assessing image quality. We associate objects with colors and inappropriate colors interfere with the HVS, e.g., a greenish skin color or a red banana do not match our experience and are thus perceived very disturbing. Moreover, missing color information may lead to loss of structure within the images, e.g., at areas with equal intensity but different hues or at smooth color gradients.

Therefore, IDMs which also take into account color information (color IDMs) are supposed to yield

better results in color-image quality assessment than the same IDMs operating only on intensity information – if color information is properly assessed. The assessment of color information is important, in particular, for color distortions such as color gamut mapping (for details see Section 2.4) or chromatic aberration. Additionally, the camera manufacturer who deals with the denoising problem would benefit from employing a color IDM as well since the perception of noise also depends on the color of the background [131].

With the advent of higher computational power and GPU programming, color IDMs will play a more important role in color-image quality assessment and therefore in image-processing tasks. Furthermore, color IDMs are supposed to show higher performance on color distortions. Nonetheless, there is still missing information about the performance of IDMs on color distortions and image quality assessment research still seeks for IDMs which highly correlate to the HVS – preferably color IDMs which allow for a more general color-image quality assessment.

1.2 Objectives

In addition to the introduction to **color-image quality assessment**, the reader will learn in this thesis about the fundamentals of an image-difference **metric** as well as about explicit examples of its **application**.

The main goal of my research is to develop a new image-difference metric – in particular a color-image-difference metric – which shows a high correlation to the human visual system on conventional distortions, such as blur, noise, and compression, as well as a competitive prediction performance on color distortions, especially color gamut mapping. Furthermore, a thorough analysis of the performance of state-of-the-art image-difference metrics on gamut-mapping distortions is presented.

The new image-difference metric is then applied to gamut mapping. The proposed metric serves as an objective function to minimize the perceived image difference between a reference image and an in-gamut image (this concept – constrained to equal-MSE images rather than in-gamut images – was also employed in Figure 1.1). Additionally, a further application of the new image-difference metric is presented: high-dynamic-range imaging.

1.3 Methodology

The following steps will be taken to achieve the research aims described in the previous section. The theoretical background which is needed to understand the proposed color image-difference metric is given in Chapter 2, especially visual psychophysics and color spaces. The next two chapters introduce objective (Chapter 3) and subjective (Chapter 4) assessment of image quality – both chapters are the foundation

for creating a new image-difference metric. Additionally, related work to existing image-difference metrics is given in Sections 3.2.1 and 3.2.2.

With this background knowledge, a first version of the proposed color-image-difference metric is developed in Chapter 5. An application of this metric is optimizing gamut mapping as described in the first part of Chapter 6. In the second part of that chapter, the optimization and thereby the metric are improved leading to the final version of the proposed color-image-difference metric. High-dynamic-range imaging as further application of color-image quality assessment is presented in Chapter 7.

Finally, the findings of this work are discussed and an outlook is given (see Chapter 8).

2

Background

The theoretical background which provides the basis for the proposed work is presented in this chapter. The topics are visual psychophysics (insight into the human visual system), visual resolution (image pixels per degree of visual angle), color spaces (quantifiable description of color), and color gamut mapping (transformation of the image's colors into the color range – called color gamut – of an output device). Please note that I do not claim that the sections are exhaustive but sufficient to understand this work.

2.1 Visual Psychophysics

This section presents a basic knowledge on the function of the *human visual system* (HVS) to better understand the perception of color-image quality. *Visual psychophysics* describes the science of interactions between visual stimuli and the sensations and perceptions of those by the HVS [38]. I.e., it quantitatively investigates how *physical* attributes of visual targets are linked to corresponding *psychological* responses [19].

Please note that in this thesis, I only present some properties of visual perception which are important to understand color-image quality. A good outline of visual psychophysics is given by Fairchild [38] who focuses on visual experiments as well as by Chandler [19] who focuses on the underpinnings of image quality. A thorough overview of psychophysics in general (including, e.g., visual and auditory psychophysics) has been presented by Gescheider [47].

2.1.1 Contrast-Sensitivity Functions

The detection of a minimum contrast of a visual target has been shown to depend on the *spatial frequency* (described in the next paragraph) of the target [19,30]. While this minimum contrast is denoted as *contrast detection threshold*, the inverse of such a threshold is denoted as *contrast sensitivity*. Functions which plot the contrast sensitivity against the spatial frequency are called *contrast-sensitivity functions* (CSFs).

Please note that in this thesis, I distinguish between threshold contrast (perceived contrast close to the contrast detection threshold), subthreshold contrast (unperceived contrast below the contrast detection threshold), and suprathreshold contrast (obviously perceived contrast far above the contrast detection threshold).

Spatial frequency SF of a visual target is a measure of how many changes of a visual item – so-called *cycles* – occur per degree of visual angle across the image. A visual item is, for instance, a single pixel on the display or a printed half-tone dot. High spatial frequencies refer to changes in close proximity, such as noise, and low spatial frequencies to changes distributed over a broad range, such as intensity shifts. The unit of spatial frequency is $[SF]$ = cycles / degree of visual angle = c/deg.

A common methodology to measure CSFs is to use sine waves which was first done by Schade [120] in 1956. In Schade's experiment, the subjects had to determine the contrast detection threshold for achromatic sine-wave gratings of different spatial frequencies. The resulting achromatic CSF was band pass with a peak contrast sensitivity near SF = 4 - 6 c/deg. However, further experiments have shown that contrast sensitivity also depends, among others, on the observers' age [152]; on the color of the gratings [91]; on the orientation of the sine-wave gratings [14]; on the visual target (e.g., Gabor functions instead of sine-wave gratings) [102]; on the adapting luminance [64].

A popular achromatic CSF which may be used for *image quality assessment* (IQA, see Chapter 3) as a prefilter [19] was proposed by Daly [28] in 1990. It is a band-pass filter which peaks at the spatial frequency SF = 1.5 - 4 c/deg depending on the adapting luminance [64]. A thorough analysis of achromatic CSFs used for IQA is given by Barten [9]. Poirson and Wandell [103,104] measured distinct CSFs for achromatic as well as for chromatic targets. For chromatic CSFs, the contrast sensitivities to red-green and blue-yellow chromatic gratings measured by Mullen [91] in 1985 have been widely used. They have the characteristics of a low-pass filter with a high-frequency cut-off at about $SF = 11$ c/deg [98] as illustrated in Figure 2.1.

2.1.2 Visual Masking

The perception of distortions can vary with the region of the image where the distortions are located. This finding is assigned to *visual masking* [72], i.e., a mask reduces the perceptibility of a given visual target [19]. To deploy visual masking in IQA, a reference image commonly serves as the mask while a distortion of the reference image serves as the visual target to be detected.

As an example, an increase in the luminance of the background – thus mask – leads to increasing detection thresholds for a target placed upon the background [12] – so-called *luminance masking*. Moreover, detection thresholds also tend to increase for increasing contrasts of masks which consist of spatial patterns [19, 72] – so-called *contrast masking*. Visual masking also applies to structural differences of

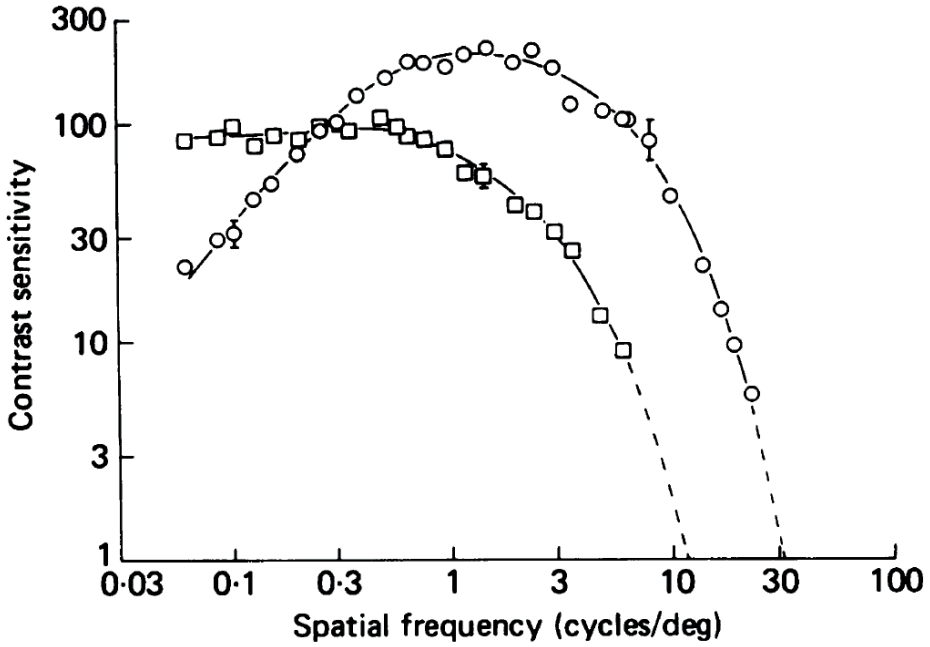


Figure 2.1: Contrast sensitivity plotted against the spatial frequency as measured by Mullen [91]. The contrast-sensitivity functions are given for a chromatic blue-yellow grating (squares; 470 nm and 577 nm) as well as for a yellow monochromatic grating (circles; 577 nm) which rather behaves like an achromatic grating. The original version of the figure has already been published in [91].

spatial patterns which is therefore called *pattern masking*. Due to the possibility to compute the mask's contrast, contrast masking plays an important role in image processing and IQA research in particular.

A demonstration of contrast masking is given in Figure 2.2 by a threshold-versus-contrast (TvC) curve. TvC curves which plot masked detection thresholds against the contrast of the mask are commonly used to illustrate the results of contrast masking. The detection threshold of the masked target generally increases with increasing contrast of the mask. However, in some cases the detection threshold decreases in a low-contrast region of the mask because mask and target may interfere depending on the dimensional relationships (see sine-wave mask) [19].

2.1.3 Multi-Channel Model of the Human Visual System

Schade used sine-wave gratings for conducting the CSF study under the assumption that any stimulus can be decomposed into single sine waves of different spatial frequency [120]. Since a square wave can be described as a superposition of sine waves, Campbell and Robson [18] measured detection thresholds for square-wave gratings as well. This experiment revealed that the HVS locally decomposes a visual stimulus into multiple spatial-frequency channels [19] – also known as the *multi-channel model* of human vision [51].

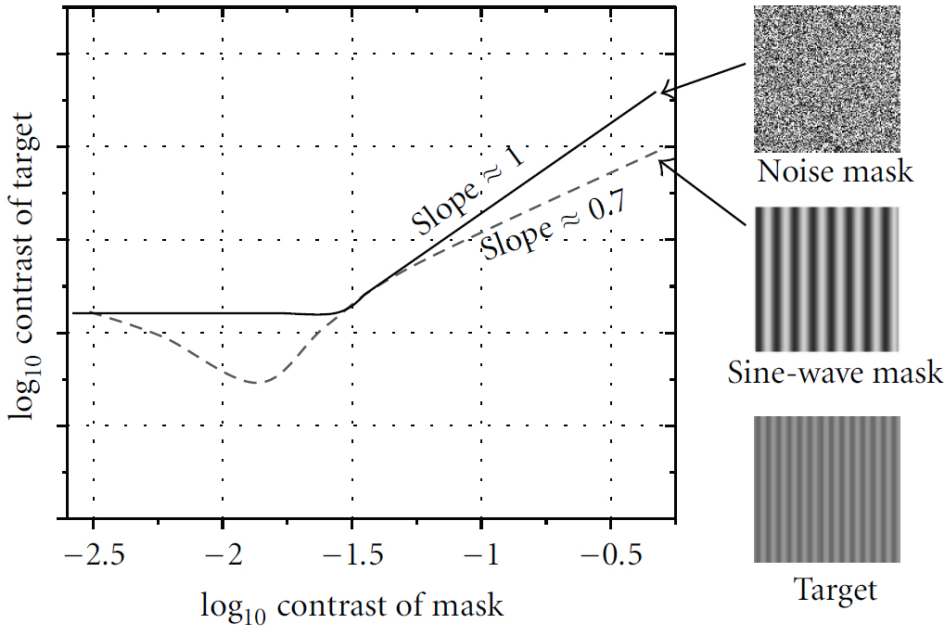


Figure 2.2: Demonstration of contrast masking for a sine-wave-grating target. The contrast detection threshold of the target is plotted against the contrast of a noise mask (solid line) and a sine-wave-grating mask (slashed line). The figure has originally been published in [19] after [72].

In image processing, the multi-channel model can be applied by regarding an image on several scales. Each scale represents the image on a different pixel resolution, i.e., the number of pixels per row and column alters while trying to keep the image content. Hence, each scale corresponds to a spatial-frequency range which furthermore simulates different viewing conditions, in particular the visual resolution (which is explained in the next section). The relative importance of each scale – needed for an appropriate assessment of image quality – is conceptually related to the CSFs of the HVS [144]. These CSFs are usually derived by measurements of threshold contrasts using sine-wave gratings. In IQA, however, suprathreshold contrasts of complex structured images are of particular importance which was taken into account by a multi-scale approach proposed by Wang et al. [151].

2.2 Visual Resolution

When a visual experiment is conducted, the viewing conditions should be constant for each subject and provided as metadata with the experimental results. Although many researchers only provide the viewing distance to the visual target as a viewing condition, rather the *visual resolution* of the target is the figure of merit. That is to say, the viewing distance is expressive only if the size of the visual target is also known. Visual resolution VR of a visual target is a measures of how many samples of a visual item –

in particular, pixels on a display – occur per degree of visual angle across the image. The unit of visual resolution is $[VR]$ = samples / degree of visual angle = s/deg.

In accordance with Section 2.1.1, visual resolution is strongly related to spatial frequency. To be more precise, if neighboring samples of a visual item are different, they form a cycle since the visual item has changed. I.e., several samples of a visual item form a cycle and cycles are therefore related to samples. On a display, a sample of a visual item corresponds to a pixel. Thereby, a minimal change – thus cycle – can be achieved by alternating two pixels. I.e., on a display 2 samples correspond to 1 cycle and therefore a visual resolution of $VR = 2\,\text{s/deg}$ corresponds to a spatial frequency of $SF = 1\,\text{c/deg}$. Please note that many authors do no distinguish between spatial frequency SF and visual resolution VR and use exclusively either the unit c/deg or s/deg.

All visual experiments that can be found in this thesis were conducted on a display. The visual resolution of an image on a display can be computed if, for instance, the viewing distance D, the number P of horizontal pixels on the screen ($[P]$ = samples = s), and the horizontal screen width W are given. The pixel pitch W/P, i.e., the pitch between two pixels, may be used instead of the horizontal screen width. Then, the visual resolution VR of samples (s) per degree of visual angle (deg) is computed by:

$$VR = P / \left[2 \cdot \arctan(W/(2D)) \cdot (180\,\text{deg}/\pi)\right]. \tag{2.1}$$

If the visual resolution is not known for a visual experiment, I suggest to assume a visual resolution $VR = 40\,\text{s/deg}$ which may refer to an office environment. For instance, a viewing distance $D = 0,75\,\text{m}$ to a screen which displays $P = 1\,200$ pixels, thus samples, on a screen width $W = 0,4\,\text{m}$ results in a visual resolution $VR = 40\,\text{s/deg}$.

2.3 Color Spaces

Intuitively, colors are described by *color names* – such as white, red, yellow, and brown – and *color adjectives* – such as light, dark, dull, and saturated [10]. However, such a description is not unambiguous for two observers, not quantifiable, depends on the viewing conditions, and hardly covers all visible colors.

Therefore, the concept of *color spaces* has been introduced which quantifies color in a color vector [117]. Each entry of the color vector refers to a coordinate of the color space representing color as a point within the color space. There are numerous color spaces designed for various purposes which may be device-dependent or device-independent with different meanings of the color-space coordinates. A detailed overview on color description and color spaces is given by Berns [10] and Reinhard et al. [117], respectively.

For the creation of a color space, some of the following properties may be desired [117]:

- **Physical realizability:** Linking the color-space coordinates with physical devices or quantities, such as cameras, displays, printers, or the cones of the HVS.
- **Intuitive color specification:** Linking the color-space coordinates with intuitive color attributes, such as hue, saturation, and lightness – thus called a *perceptual color space*. Ideally, the coordinates are free from *cross contamination*, i.e., a change in one color attribute does not affect the other perceived color attributes.
- **Perceptual uniformity:** The Euclidean distance of two colors in the color space corresponds to the perceived color difference.
- **Efficient encoding:** Storing color images in a computationally inexpensive representation without loss of color accuracy.

The concept of color spaces allows to define so-called *color-difference formulas* – the quantitative distance between two colors in the color space. Such a distance – ideally equivalent to a perceived color difference – simplifies the comparison of color differences independently of the colors' position in the space [153]. A simple color-difference formula for a color space with Cartesian coordinates is the Euclidean distance. However, to account for shortcomings of a specific color space (for instance, perceptual non-uniformity), more complex color-difference formulas may be defined on that color space.

In the subsequent sections, only those color spaces are presented which have been used in this work. Specific transformations between color spaces can be found in the Appendix A.2. A more thorough overview of different color spaces is given in [117, 153].

2.3.1 RGB Color Spaces

The RGB (*red-green-blue*) color model is an additive color model, i.e., colors are defined as an additive mixture of the three primaries red, green, and blue [54]. Zero intensity of all primaries refers to black and full intensity of all primaries to white. The model is derived from color perception of the HVS in which the LMS (*long-medium-short* which refers to the wavelength of peak sensitivity) cones of the human eye are involved.

Most RGB color spaces are based on the RGB color model and are often defined by the *chromaticities* of the three primaries and the *white point* as well as by a *gamma compression* of the primaries. In digital imaging, the most common way to represent a color image for storing in an image file is an RGB image – a representation in a trichromatic RGB color space. For instance, digital cameras write out captured images as RGB images corresponding to their RGB sensor and monitors read in RGB images corresponding to their RGB primaries. For the processing of color images, the color images are usually

represented in a so-called *working color space* which should allow an easy access to color attributes, e.g., lightness and hue. However, RGB color spaces are usually less feasible as working color space.

In this thesis, sRGB – *standardized RGB* color space – is exclusively used as the RGB color space present in the image file of the color images. The specification of the sRGB color space is given in the Appendix A.1. This color space is the most commonly used RGB color space in image processing. Other common RGB color spaces including, among others, *Adobe RGB* or *eciRGB* could be used as well as color space for the employed color images.

2.3.2 CIEXYZ

The *CIEXYZ color space* was introduced by the *CIE* – the *International Commission on Illumination* (*Commission internationale de l'éclairage*) – in 1931. The goal was to create a device independent color space which includes all real colors and is based on the *photopic luminous efficiency function* $V(\lambda)$. The primaries of the CIEXYZ color space are imaginary, i.e., they are not physically realizable colors [153]. However, to construct such a color space imaginary primaries are required. CIEXYZ values are represented as non-negative tristimulus values $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$.

2.3.3 CIELAB

The disadvantage of the CIEXYZ color space is that it is not perceptually uniform and that the tristimulus values $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ can hardly be interpreted. RGB color spaces are not perceptually uniform as well. Influenced by the opponent color theory and the Munsell color system, the CIE proposed the *CIELAB color space* in 1976. The perceptual CIELAB color space is separated into an achromatic channel – *lightness* L^* – and two chromatic channels – *red-green* a^* and *blue-yellow* b^*.

The chromatic channels a^* and b^* can be represented in polar coordinates resulting in more meaningful coordinates. The radius corresponds to the chroma predictor C^* and the angle corresponds to the hue predictor H^* of the color (see Equations A.20 and A.21 in the appendix). Thus, lightness L^*, chroma C^*, and hue H^* meaningfully describe a color relative to a given white point.

2.3.4 LAB2000HL

The CIELAB color space is the most popular opponent color space, it is perceptually more uniform than the CIEXYZ color space and it has meaningful coordinates – *lightness* L^*, *red-green* a^*, and *blue-yellow* b^* and *lightness* L^*, *chroma* C^*, and *hue* H^*, respectively. Nonetheless, the CIELAB color space is not a perfect color space since its perceptual uniformity is not sufficient for highly accurate applications like proofing and it is not hue linear, i.e., a path of perceptually constant hue in the a^*b^*-plane is not a straight line.

A perfect perceptually uniform and hue linear color space does not exist [77] but attempts were made to fulfill at least one property, e.g., the hue linear IPT color space [32]. A hue linear and nearly perceptually uniform color space was introduced by Lissner and Urban [78] – the perceptual LAB2000HL color space. It is derived from the CIELAB color space by forcing hue linearity while reducing perceptually non-uniformity. Accordingly, its coordinates are called *lightness* L, *red-green* a, and *blue-yellow* b. The representation of the chromatic channels in polar coordinates *chroma* C and *hue* H is done analogously to the CIELAB color space.

2.4 Color Gamut Mapping

The so-called *color gamut* of a device (e.g., printer or display) is the set of all colors reproducible by this device for a given set of viewing condition. Every device has a limited color gamut which requires a transformation if colors of an image to reproduce are outside the gamut. This transformation is referred to as *color gamut mapping*, or simply *gamut mapping*. A detailed overview of this topic is given by Morovič [88].

A color gamut can be represented within a color space as a complete subset of colors. As a subset of a color space, the color gamut can also change its representation by color-space transformations. This allows an easier access to gamut mapping, especially if the image and the gamut of the device are represented in the same color space, e.g., the CIELAB color space [16]. Gamut mapping which is a rather complex task is generally performed by so-called *gamut-mapping algorithms* (GMAs) which might be encoded in *look-up tables* if the characteristics of the devices are known.

3

Image-Difference Metrics

This section gives an introduction to image quality and how image quality is assessed. As image quality assessment is mostly a relative-to-reference evaluation, the more general *image-difference metric* will be defined – an automatic comparison of two arbitrary images of the same resolution. Image-difference metrics are represented by algorithms which automatically assess the perceived difference of two images. Related work to image-difference metrics is given as well as a detailed description of metrics employed in this thesis.

3.1 Image Quality

Image quality has been attracted interest since the invention of optical instruments which dates back to the beginning of the 17^{th} century [35]. More interest in image quality emerged with the introduction of photography and television [99]. Today, image quality plays an active role in image processing, e.g., in applications such as image and video encoding, digital watermarking, denoising, or image synthesis [19].

From the various recent definitions of image quality – e.g., proposed by Engeldrum [36], Jacobson [57], or Keelan [62] – I present a definition from the *International Imaging Industry Association (I3A)* [56] in good accordance with my usage of image quality in this thesis:

> *"Image quality is the perceptually weighted combination of all visually significant attributes of an image when considered in its marketplace or application."* [56].

I.e., image quality ...

- ... is a subjective impression and perceived differently for each observer.
- ... consists of visually significant attributes – in this thesis called *features* – which are weighted and combined by the *human visual system* (HVS).

- ... depends on its context and application. For instance, when I process a digital photograph to get a higher "quality", I could aim to recover the original scene or to enhance the contrast for a more colorful image.

As image quality depends on the application, it is a relative rather than an absolute perception. In some cases, the highest image quality refers to real-world scenes even if the scene cannot be depicted by digital images. In some other cases when an image needs to be processed (e.g., image compression or gamut mapping), the highest quality refers to the starting image called *reference image*. The processed images which are called *distorted images* are often intended to be as similar to the reference image as possible.

3.1.1 Image Quality Assessments

In image processing, we are interested in extracting the perceived quality out of a digital image. The most natural way to obtain image quality is to ask an observer about his judgment. Since each observer judges image quality differently, a lot of observers need to be asked in a visual experiment with consistent viewing conditions for an estimate of the image quality (see Section 4.1). Although the results of a well-designed visual experiment can be considered as ground truth of image quality, such a procedure is exhausting, time-consuming and therefore expensive. Moreover, it is not feasible in most applications of image processing, such as image transmission or real-time machine vision.

This is where *image quality assessment* (IQA) comes into play. IQA is an objective and quantitative measure which automatically predicts perceived image quality [149]. Over the last several decades, numerous algorithms for IQA have been investigated and proposed. Since these algorithms play an important role in image-processing applications today, "IQA research has emerged as an active subdiscipline of image processing" [19]. Depending on the availability of information provided to assess image quality, IQA can be divided into three subgroups:

- **Full-reference IQA:** The quality of processed (usually distorted) images with respect to the reference image of which they originate is predicted by full-reference IQA. This is a relative IQA which generally assigns the best quality to a processed image that is perceptually equal to the original image.
- **No-reference IQA:** If a reference image is not available, the perceived quality of an image is predicted by no-reference IQAs. This is an absolute IQA which usually detects specific distortion types.
- **Reduced-reference IQA:** In some cases, there is no reference image available but partial information about the reference images is provided. Reduced-reference IQA employs this side information to evaluate distorted images.

Mostly full-reference IQA is used in image processing. The latter two IQAs, however, occur in many practical applications [149]. But they are out of the scope and not used in this thesis which only refers to full-reference IQA. For the interested readers, a brief survey of no-reference and reduced-reference IQA algorithms has been presented by Chandler [19].

3.2 Introduction to Image-Difference Metrics

In full-reference IQA, a distorted image is compared to its reference. In this thesis, I even propose a more general approach: the comparison of two arbitrary images $\mathbf{X}, \mathbf{Y} \in \mathbb{I}_{M,N}$, where $\mathbb{I}_{M,N}$ is the set of all sRGB images with M rows and N columns. Please note that any colorimetric color space can be used for representing color images even though I restrict to sRGB images stored in the images' file format without loss of generality (see Section 2.3.1). Moreover, in this work only pairs of images with the same image resolution are considered for image comparisons.

This approach demands a different denomination as we are interested in the perceived difference of two images rather than the relative-to-reference quality. Therefore, I define the comparison of two images $\mathbf{X}, \mathbf{Y} \in \mathbb{I}_{M,N}$ as *image-difference metric* (IDM) $\mathbf{d}(\mathbf{X}, \mathbf{Y})$ with the following property:

$$\text{Property 1: } \mathbf{d}(\mathbf{X}, \mathbf{Y}) \geq 0 \qquad \text{(non-negative)}, \tag{3.1}$$

i.e., $\mathbf{d}(\mathbf{X}, \mathbf{Y})$ returns a single non-negative number for the perceived image difference of $\mathbf{X}, \mathbf{Y}$. Full-reference IQAs are thus a special case of IDMs. Please note that in literature, the denomination *image-difference metric* is not exactly defined and IDM is used in the same way as full-reference IQA [52].

The term *metric* which I employ in the denomination *image-difference metric* refers to an objective assessment of image differences. Originally, the term *metric* refers to a mathematical formulation for measuring a distance. The correct mathematical formulation of a metric $\mathbf{d}(\mathbf{X}, \mathbf{Y})$ on a set $\mathbb{I}_{M,N}$ is defined by the subsequent properties $\forall\, \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{I}_{M,N}$ [73]:

$$\text{Property 2: } \mathbf{d}(\mathbf{X}, \mathbf{Y}) = \mathbf{d}(\mathbf{Y}, \mathbf{X}) \qquad \text{(symmetry)} \tag{3.2}$$

$$\text{Property 3: } \mathbf{d}(\mathbf{X}, \mathbf{Y}) = 0 \quad \Leftrightarrow \quad \mathbf{X} = \mathbf{Y} \qquad \text{(identity of indiscernibles)} \tag{3.3}$$

$$\text{Property 4: } \mathbf{d}(\mathbf{X}, \mathbf{Z}) \leq \mathbf{d}(\mathbf{X}, \mathbf{Y}) + \mathbf{d}(\mathbf{Y}, \mathbf{Z}) \qquad \text{(triangle inequality)}. \tag{3.4}$$

These three properties imply Property 1 which is why in many cases Property 1 is also listed as a property for a metric.

The least property an IDM must fulfill is Property 1 but Properties 2 - 4 are also desirable. Symmetry (Property 2) ought to be valid for an IDM but not necessarily for a full-reference IQA which compares

two distinct images (reference and distorted image). In image processing, the property of symmetry is generally fulfilled for both full-reference IQAs and IDMs. Property 3 only holds for identical images. However, the HVS may perceive two distinct images as equal even though they are not identical – e.g., the deviation of a pixel's color is below the perceptual threshold. An IDM should reflect this by returning zero if two images are perceived equal even though they are not identical. Nevertheless, the vast majority of IDMs fulfill Property 3 only for identical images but not for perceptually equal images. Please note that some IDMs do not return zero for a pair of identical images, e.g., $\mathbf{d}(\mathbf{X}, \mathbf{X}) = 1$ for IDMs analog to correlation coefficients. The triangle inequality (Property 4) is not applicable for perceived image differences. For instance, an image $\mathbf{X}$ can be changed to an image $\mathbf{Y}$ in such a way that the change is below perceptual threshold and the IDM detects perceptual equality, i.e., $\mathbf{d}(\mathbf{X}, \mathbf{Y}) = 0$. If $\mathbf{Y}$ is changed again below threshold to an image $\mathbf{Z}$ which has a visible difference to $\mathbf{X}$, the triangle inequality is not valid. Therefore, the property of triangle inequality is omitted for defining the assessment of perceived image differences. Instead, the desired property of an IDM is that the metric returns a larger value if the perceived image difference is bigger (for more details see Equation 6.1 later in this thesis).

To summarize, an image-difference metric (IDM) assesses the perceived image difference and expresses the assessment as a single non-negative number. Further desired properties are symmetry, identity of indiscernibles for equally perceived images, and return of larger values for bigger perceived image differences. Not mentioned in this section is that IDMs are depending on the viewing conditions. To also account for the viewing conditions, the definition of IDM has to be expanded which is done later in Section 5.3. If the IDM furthermore accounts for the viewing conditions described by a parameter array $\mathbf{P} \in \mathcal{P}$ where $\mathcal{P}$ is the set of all parameter arrays, the mathematical formulation of the IDM's prediction is referred to as $\mathbf{IDM}(\mathbf{X}, \mathbf{Y}, \mathbf{P})$.

In the color research community, the term *full-reference image quality assessment* is widely used for the objective evaluation of perceived image differences [19,149,155]. Moreover, the image-processing community does not distinguish between full-reference image quality assessment (comparison of a distorted to its reference image) and image-difference metric (comparison of two arbitrary images) – both are only different terms for the same process: comparison of two images. Further denominations for the comparison of two images used by some authors are, for instance, index [149], measure [93], or fidelity [127]. In the following, I use the more general term *image-difference metric* (IDM) for algorithms which evaluate the perceived quality between two images even though only relative-to-reference IQAs occurs.

3.2.1 Existing Image-Difference Metrics

There are various IDMs which have been published for different purposes, such as quantifying distortions, benchmarking, monitoring quality, or optimizing processes [99]. It is unlikely that there is an IDM which is feasible for all possible applications and furthermore shows a high performance on them. Hence, it

is important to know the context of IDMs to find an appropriate metric for a given applications. A classification of IDMs according to their attributes as done by Pedersen and Hardeberg [101] may be useful.

Existing IDMs are based on a diversity of approaches and some of these approaches are quite similar. Therefore, various researchers have categorized IDMs into groups for an easier overview. For instance, to name only a few authors, Avcıbaş et al. [5] divided IDMs into six categories, Chandler [19] divided IDMs into four categories, Pedersen and Hardeberg [101] divided IDMs into eight categories, and Seshadrinathan and Bovik [125] as well as Thung and Raveendran [140] divided IDMs into three categories. Please note that it is not always possible to classify an IDM into a single category [147].

I propose to divide IDMs into four categories subject to the approach they are based on:

- **Noise:** Assessment of noise and pixel deviations of the images, e.g., *peak signal-to-noise ratio* (see Section 3.3.1). Such IDMs regard a distortion as a pixel-wise deviation from the reference image which in turn refers to noise. These pixel-wise deviations are usually weighted by models of the HVS.

- **Structure:** Assessment of local structural deviations of the images, e.g., *Structural Similarity* (see Section 3.3.2) and *Feature Similarity* (see Section 3.3.4). Since the HVS is sensitive to changes in contrast and structure (a more detailed explanation will be given in Section 5.3) which cannot be detected by regarding single pixels, such IDMs assess image distortions within local areas.

- **Statistics:** Assessment of statistical deviations of the images, e.g., *Visual Information Fidelity* (see Section 3.3.5). Images exhibit statistical properties from which information content is extracted and then assessed by such IDMs.

- **Other techniques** Approaches which do not fit into the preceding categories.

The fourth category of other techniques is reasonable [19] but is not regarded in the following because all IDMs which appear in this work fit into the other categories.

From the various IDMs proposed so far, I only present a small selection in Section 3.3. However, this selection contains important IDMs which are well-known in IQA research. Moreover, the presented IDMs – also representing the categories of noise, structure, and statistics – are explained in detail to give the reader an understanding of the design of IDMs.

For more than 50 years, the standard metric in signal processing is the *mean-squared error* due to its simple structure and its fast computation [148]. The shortcoming is that it totally disregards the content of visual scenes and therefore shows a low correlation to human perception of image quality. Thus, IDMs are supposed to assess the images' content, e.g., by employing models of the HVS, detecting contrasts and structures, or employing image statistics. Today, an important state-of-the-art IDM in image processing is

the *Structural Similarity* index proposed by Wang et al. [149] in 2004. It extracts structural information from the images and compares them locally to derive an image-difference prediction. Up to now, the best performing IDM on the largest publicly available IQA database is the color version of the *Feature Similarity* index proposed by Zhang et al. [155] in 2011 which is based on the Structural Similarity index. A rather different approach has been proposed by Sheikh and Bovik [127] in 2006. They employed *natural scene statistics* to detect deviations in the distorted image with respect to the reference image. Then, these deviations are quantified by an image-information measure and combined into relative image information called *Visual Information Fidelity*.

In addition to the mentioned IDMs, numerous metrics were developed in the last decades. Several authors summarized a large choice of IDMs into a survey – I recommend the surveys given in the following for an overview of existing IDMs. Chandler [19] presented a review of full-reference IQAs as well as of reduced-reference and no-reference IQAs in 2013. An extensive survey of more than 100 IDMs have been recently proposed by Pedersen and Hardeberg [101] in 2012 who furthermore classified the IDMs according to their scope of use. Other interesting surveys were given by Avcıbaş et al. [5] in 2002 covering statistical IDMs and by Wang and Bovik [147] in 2006 covering IDMs which mainly assess grayscale images. An overview of IDMs operating on color images is given in the next section.

3.2.2 Color in Image-Difference Metrics

The vast number of image-processing applications consider color images since a real-world scene is usually perceived in color. Color images correspond in this work to images represented in the sRGB color space – which is predominantly used in digital imaging – even though any color space present in color-image file formats could be used. sRGB images possess three color channels in contrast to grayscale images which only possess one channel. A transformation from sRGB images to grayscale images is possible (see Section A.2.1 in the appendix) maintaining structure and contrast but losing, of course, the chromatic information. To distinguish between color and grayscale images in this thesis, I denote sRGB color images as $\mathbf{X}, \mathbf{Y}$ and grayscale images as X, Y.

Despite the lower information content of grayscale images, they are quite interesting in image-processing tasks. For instance, less channels and thus less complexity are beneficial for applications which need fast computation and for solving optimization problems. The same holds for assessing image quality – there are color IDMs $\mathbf{d}(\mathbf{X}, \mathbf{Y})$ operating on color images and grayscale IDMs $\mathbf{d}(X, Y)$ operating on grayscale images.

The majority of IDMs are operating on grayscale images. The survey of IDMs presented by Pedersen and Hardeberg [101], for instance, presents 75 grayscale IDMs in comparison to 59 color IDMs. Moreover, employing grayscale IDMs instead of color IDMs is justified for many applications in image processing. In particular, on the *LIVE IQA database* (introduced in Section 4.3.1) which includes common distortions

such as blur, noise, or compression, the best performing grayscale IDM (VIF, introduced in Section 3.3.5) is just as good as the best performing color IDM (in particular FSIMc, introduced in Section 3.3.4) [155].

However, it is obvious that predictions of grayscale IDMs do not highly correlate to color distortions, such as hue shifts, because they cannot detect chroma and hue deviations. Color IDMs, on the contrary, are supposed to yield higher performance when assessing color distortions.

A methodology of creating color IDMs is to employ color-difference formulas (see Section 2.3). A color-difference formula predicts the perceived color difference of two colors in the color space, e.g., the Euclidean distance $\Delta E^*_{ab}(\mathrm{x},\mathrm{y}) = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$ where x,y are colors represented in the CIELAB color space. Color IDMs based on color-difference formulas are obtained by averaging the color difference of all image pixels, e.g., $\overline{\Delta E^*_{ab}}(\mathbf{X},\mathbf{Y}) = \frac{1}{M \cdot N} \sum_{i=1}^{M \cdot N} \Delta E^*_{ab}(\mathrm{x}_i, \mathrm{y}_i)$ where $\mathbf{X},\mathbf{Y}$ are color images with M rows and N columns and $\mathrm{x}_i, \mathrm{y}_i$ are the corresponding colors in CIELAB coordinates at pixel position i, $i \in 1, ..., M \cdot N$.

Nevertheless, such pixel-based IDMs do not highly correlate with human judgments as they neglect the image content, such as contrast and structure. Instead, spatial extensions of the color-difference formulas may be employed to account for the image content [58], such as *S-CIELAB* proposed by Zhang and Wandell [156] in 1997. Recent color IDMs which employ color-difference formulas with spatial extensions include, among others, the *Spatial* ΔE_{00} proposed by Chen et al. [21] in 2008 which employs the ΔE_{00} color-difference formula [79] in combination with the *contrast-sensitivity functions* (CSFs) and the *Cortex Transform decomposition* extracted from the *Visible Difference Predictor* proposed by Daly [27]; the *Spatial-DEE* proposed by Simone et al. [132] in 2009 which employs the ΔE_E color-difference formula [95] in combination with an extension of S-CIELAB; the S_{DOG}*-CIELAB* proposed by Ajagamelle et al. [1] in 2010 which employs the ΔE^*_{ab} color-difference formula in combination with an extension of S-CIELAB and the *difference of Gaussians receptive-field model* [138].

A common strategy to develop color IDMs is to extend existing grayscale IDMs to color. This can be done, for instance, by combining the predictions of the grayscale IDM computed on each color channel or by adding color-assessment functions to the grayscale IDM. Examples of such color extensions are the *Colour Image Fidelity* metric Q_{color} proposed by Toet and Lucassen [142] in 2003 which is based on the *Universal Image Quality Index* [146] and computed on transformations of the *LMS* color space [119]; *SSIM-IPT* proposed by Bonnier et al. [13] in 2006 which is based on the *Structural Similarity* index [149] and computed on the *IPT* color channels [32]; χ*SSIM* proposed by Scheller Lichtenauer et al. [121] in 2012 who added features for hue and chroma shifts to the Structural Similarity index.

3.3 Selection of Image-Difference Metrics

More than 100 IDMs have been proposed in literature [100]. However, in the upcoming section only a small fraction of important IDMs is presented which are also used in this thesis. Thereby, IDMs of each category of noise, structure, and statistics (see previous section) are represented.

3.3.1 Mean-Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR)

Wang and Bovik stated that "for more than 50 years, the mean-squared error (MSE) has been the dominant quantitative performance metric in the field of signal processing" [148]. Its simple and intuitive form makes it suitable for use in image and signal processing. The MSE between two grayscale images X and Y is

$$\mathrm{MSE}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2, \tag{3.5}$$

where N is the number of the image's pixels and X_i and Y_i are the corresponding pixel values of pixel i.

In many cases, it is rather usual to refer to the *peak signal-to-noise ratio* (PSNR). The PSNR is deduced from the MSE but also includes information about the dynamic range DR of the image (here, number of possible image-pixel intensities, e.g., $DR = 255$ for an 8-bit image [148]) for a better comparison of metrics from different dynamic ranges:

$$\mathrm{PSNR}(X,Y) = 10\log_{10}\frac{DR^2}{\mathrm{MSE}(X,Y)}. \tag{3.6}$$

MSE and PSNR evaluate image differences based on pixel deviations (which refers to noise in terms of signal processing). This allows an easy and quick computation making them the right choice for some image-processing applications. However, their correlation to human perception is low [147] because they do not consider the content of the images. For instance, the same distortion on a uniform area is perceived more disturbing than on a textured area but leads to the same prediction. More complex models of signal-to-noise-ratio approaches which also account for the HVS are supposed to yield better results, e.g., the *Visual Signal-to-Noise Ratio* (VSRN) [20]. A concept different from these pixel-based IDMs (category of noise) – in particular structure-based IDMs (category of structure) – is introduced in the upcoming sections.

3.3.2 Structural Similarity (SSIM) Index

The *Structural Similarity* (SSIM) index [149] is a renowned IDM which assesses the images locally with good prediction performance on IQA databases, e.g., on the TID2008 [109] or LIVE [129] databases (see Section 4.3). It extracts local image-difference comparisons – luminance comparison l, contrast comparison c, and structure comparison s – which are combined resulting in the metric. The SSIM index

is an extension of the *Universal Image Quality Index* (UQI) [146]. As an input, the SSIM index takes grayscale images X,Y.

First, local statistics are computed on the images – the mean μ_x, the standard deviation σ_x, and the covariance σ_{xy} corresponding to luminance, contrast, and structure, respectively. They are computed within a $k \times k$ square window. Please note that k has to be odd. The local window pixels of images X and Y are written into corresponding pixel arrays x and y which have the length $m = k^2$. Statistics are based on a normalized $k \times k$ circular-symmetric Gaussian weighting $w = \{w_i, i = 1, ..., m | \sum_{i=1}^{m} w_i = 1\}$ with standard deviation σ_G:

$$\mu_x = \sum_{i=1}^{m} w_i x_i \tag{3.7}$$

$$\sigma_x = \left(\sum_{i=1}^{m} w_i (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \tag{3.8}$$

$$\sigma_{xy} = \sum_{i=1}^{m} w_i (x_i - \mu_x)(y_i - \mu_y). \tag{3.9}$$

Then, the local statistics are used to extract the image-difference comparisons for pairs of corresponding pixel arrays (x, y):

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{{\mu_x}^2 + {\mu_y}^2 + c_1} \tag{3.10}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{{\sigma_x}^2 + {\sigma_y}^2 + c_2} \tag{3.11}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{3.12}$$

where $c_1, c_2, c_3 > 0$ are constants accounting for the images' representation.

Image-difference comparisons computed on the whole image can be visualized by so-called *image-difference maps*. In an image-difference map no difference is indicated by black pixels and larger differences by brighter pixels (examples of image-difference maps are given Figure 5.4 and 5.6 in Chapter 5).

Finally, the comparisons are combined using a factorial model and then pooled by taking the mean:

$$\mathrm{SSIM}(X,Y) = \overline{l(x,y)^{\alpha_1} \cdot c(x,y)^{\alpha_2} \cdot s(x,y)^{\alpha_3}} \tag{3.13}$$

where $\alpha_1, \alpha_2, \alpha_3 > 0$ are exponents to weight the comparisons due to their importance and $\overline{f(x,y)}$ denotes the mean over all pairs of corresponding pixel arrays (x, y) within the images X, Y. Contrary to the definition of IDMs, the SSIM index assigns no image difference to value 1 and larger differences to smaller values.

In practice, the SSIM index is used with predefined parameters. They are summarized in Table 3.1. In the thesis, I refer to these values when using SSIM.

Table 3.1: Parameters of the SSIM Index.

c_1	c_2	c_3	α_1	α_2	α_3	k	σ_G
6.5025	58.5225	29.26125	1	1	1	11	1.5

Table 3.2: Scale Weights of the MSSIM Index.

β_1	β_2	β_3	β_4	β_5
0.0448	0.2856	0.3001	0.2363	0.1333

3.3.3 Multi-Scale SSIM (MSSIM) Index

The SSIM index presented in the previous section is single-scale IDM, i.e., the image-difference comparisons luminance l, contrast c, and structure s are computed on a single spatial frequency band. Wang et al. [151] proposed a multi-scale approach which compares the image details at different scales, thus called *Multi-Scale Structural Similarity Index* (MSSIM index). It takes into account that the *human visual system* (HVS) has different sensitivities across scales (see Section 2.1.3).

The MSSIM index has also two grayscale images X, Y as input. To compute the next higher scale both images are low-pass filtered and then downsampled by a factor of 2. A total of m scales are used starting with scale 1 for the original size. The image-difference comparisons on the i-th scale ($i \in \{1, ..., m\}$) are denoted as $l^i(x, y)$, $c^i(x, y)$, and $s^i(x, y)$, respectively. To calculate the MSSIM index the local image-difference comparisons are averaged for each scale and then multiplied:

$$\mathrm{MSSIM}(X, Y) = \overline{l^m(x, y)}^{\beta_m} \cdot \prod_{i=1}^{m} \left(\overline{c^i(x, y)} \cdot \overline{s^i(x, y)} \right)^{\beta_i} \tag{3.14}$$

where $\beta_1, ..., \beta_m > 0$ adjust the corresponding scale and $\overline{f(x, y)^i}$, $i \in \{1, ..., m\}$ denotes the mean over all pairs of corresponding pixel arrays (x, y) within the images X, Y on the i-th scale.

The parameters $\beta_1, ..., \beta_m$ are normalized such that $\sum_{i=1}^{m} \beta_i = 1$. Hence, they are weighting each scale due to their importance. They might be related to *contrast-sensitivity functions* (CSFs, see Section 2.1.1) which detect the visibility threshold of simple stimuli. However, since complex-structured images rather than simple stimuli are considered, an investigation of the scale weights was done by Wang et al. [151]. They employed an image synthesis approach to generate images associated with specific distortion levels on different scales assuming a visual resolution of $VR = 32\,\mathrm{s/deg}$. In a visual experiment, the different scales were compared with respect to the quality of the distorted images for $m = 5$ scales. The resulting scale weights of the MSSIM index to which I refer in the thesis are given in Table 3.2.

3.3.4 Feature Similarity Index (FSIM)

The concept of assessing local structure instead of pixel deviations in IQA was employed very successfully by the SSIM and MSSIM index. They yielded the best prediction performance for a selection of state-of-the-art IDMs [109] on the renowned *Tampere Image Database 2008* (TID2008, see Section 4.3.2). Inspired by SSIM, Zhang et al. [155] developed an IDM in the year 2011 based on the similarity of low-level features, thus called *Feature Similarity* (FSIM) index. FSIM operates only on grayscale images – for color images, they proposed the *FSIMc* index which additionally incorporates chrominance information.

According to Zhang et al. [155], low-level features are features of an image that are crucial for the HVS to interpret the scene, such as edges or zero-crossings. FSIM extracts two features, the *phase congruency* (PC) and the *gradient magnitude* (GM). While the contrast invariant PC measures the significance of a local structure, the complementary GM encodes contrast information. Applying PC and GM on the whole image results in so-called *local-quality maps* which are then pooled (PC-weighted mean) to obtain a single value, the image-difference prediction. To determine the unknown parameters, FSIM was trained on a part of the TID2008 by maximizing the Spearman rank-order correlation [135].

The performance of FSIM was tested on several state-of-the-art IQA databases [155]. In all cases, FSIMc performs better than FSIM. Furthermore, it yielded the best correlation to human perception among state-of-the-art IDMs on almost all IQA databases. Moreover, FSIMc showed the best performance on the largest publicly available IQA database, the *Tampere Image Database 2013* (TID2013, see Section 4.3.2) [107]. Please note that TID2008 is integrated in TID2013, i.e., FSIM is thus trained on a part of TID2013 which does not allow a fair comparison to IDMs not trained on the database.

However, these results were verified only on conventional distortions (see Section 4.3) such as noise or blur. The performance of FSIM on gamut-mapping distortions (see Section 4.4) is given later in Section 5.5.4.

3.3.5 Visual Information Fidelity (VIF)

In addition to IDMs based on assessing noise (such as PSNR) or local structure (such as SSIM), I introduce the concept of the third category of IDMs: statistics-based IDMs. Such metrics employ the statistical properties of images, in particular of natural images which broadly correspond to captured real world scenes [128].

An established representative of IDMs based on statistics is the *Visual Information Fidelity* (VIF) proposed by Sheikh and Bovik [127] which is defined only for grayscale images. This metric is based on natural scene statistics which have already been used in various image-processing algorithms like compression, denoising, and texture analysis. The reference image is considered as stochastic natural image source from which information content is extracted by a statistical model corresponding to the

output of the HVS. The distorted image is considered as being an alternation of the natural image source which deviates from natural statistics. Then, an image-information measure quantifies the information content of the reference image and the loss of information in the distorted image. Finally, VIF combines both measures to relative image information which is strongly related to visual quality.

On the renowned TID2008 IQA database (presented in Section 4.3.2), VIF shows a good correlation to human perception and outperforms most IDMs [105]. The quite interesting statistical approach shows a high potential for designing well-performing IDMs.

3.3.6 MeTriX MuX Visual Quality Assessment Package

A publicly available selection of state-of-the-art IDMs was proposed by Gaubatz [46]. He combined twelve IDMs into the *MeTriX MuX Visual Quality Assessment Package* which provides MATLAB code to compute their predictions of image differences. This package contains the default SSIM and MSSIM, the *Universal Image Quality Index* (UQI) [146], VIF and the *pixel domain version of VIF* (VIFP) [127], the *Information Fidelity Criterion* (IFC) [128], the *Noise Quality Measure* (NQM) [29], signal-to-noise ratio (SNR), the *Visual SNR* (VSRN) [20], the *Weighted SNR* (WSNR) [85], peak signal-to-noise ratio (PSNR), and mean-squared error (MSE).

To conclude, *objective* image quality assessment was introduced in this chapter. In particular, image-difference metrics which predict perceived image differences were addressed. Even though many image-difference metrics are based on models of the human visual system, their predictions do not have to show a high correlation to human judgments. To test the performance of image-difference metrics, *subjective* assessment of image quality is presented in the next chapter.

4

Image-Quality-Assessment Databases

The purpose of this chapter is an investigation of *image-quality-assessment databases* which are needed in image processing. The choice of images, generation of distortions, and psychophysical experiments to get information about perceived image quality are discussed. I divide image-quality-assessment databases into groups of conventional distortions such as noise and compression, and of gamut-mapping distortions which mainly include color distortions. The most renowned databases in the color-science community are presented.

4.1 General Properties of Image-Quality-Assessment Databases

The actual quality of an image refers to the perception of the majority of human observers – in contradiction to *image quality assessment* (IQA) which predicts image quality automatically (IQA is described in Chapter 3). Even though every single human has a different comprehension of quality, most humans coincide with the assignment of a certain quality for a given example. *Image-quality-assessment databases* (IQA databases) which represent the ground truth of image quality are important for various applications:

- Investigation of relationship between visual stimuli and responses of the *human visual system* (HVS) [19].
- Testing and comparing the performance of different IQAs.
- Adjustment or training of parameters of IQAs.

The task of IQA databases is to collect a lot of human assessments for a variety of images and alternations from those images which are called *distortions*. The choice of images and distortions depends on the purpose of the IQA database. In most cases, it makes sense to generate a database covering a diversity of scenes. There are natural scenes with landscapes, portraits, animals, buildings, sports, etc.

as well as artificial images like paintings, sketches, animated images, etc. The more different scenes occur in the database the more is the information value.

Every observer shows a variability in his choices even for the same comparison, e.g., I prefer image **X** to **Y** in one trial but later on I prefer **Y** to **X** (not remembering what I preferred before). This is called *intra-observer variability*. In addition to that, every observer has its own preferences in image quality, e.g., one prefers brighter images and another prefers darker images. Summarizing the results of the same experiment for all observers also shows a variability, the so-called *inter-observer variability*. Thus, an IQA database reflects the inter- and intra-observer variability which, however, can only be determined with some uncertainty. Due to statistical reasons, a large number of observers and evaluations reduces the uncertainty of an IQA database. However, generating large image-quality-assessment databases is time-consuming, exhaustive, and therefore expensive. A good balance between number of observers and comparisons on the one hand and efficiency on the other hand is the figure of merit.

There are many possibilities to gain information about image quality. A good overview about visual psychophysics (see Section 2.1) and experimental design is given by Fairchild [38]. A modern and thorough review of psychometric scaling as well as its techniques and applications can be found in [36]. Visual experiments are divided into two groups: *threshold and matching experiments* on the one hand and *scaling experiments* on the other hand.

In threshold and matching experiments the *just-noticeable difference* (JND) is investigated which is the threshold of detecting a change in a stimulus. This refers to the observers' sensitivity of a given stimulus [38]. While threshold experiments investigate a change in a stimulus, matching experiments ask for a match between stimuli. The problem is not easy due to inter- and intra-observer variability, i.e., the absolute threshold does not remain constant for every observer and measurement. The JND also depends on the experimental method, e.g., going stepwise from clearly perceptible difference to non-perceptible difference normally results in a different JND than going the reverse direction. The methods of adjustment, limits, and constant stimuli are also worth mentioning which are well-described in [38].

Mostly used for IQA databases are scaling experiments which specify the relationships between stimuli [38]. There are four groups of scales. The easiest scales are nominal scales which simply separate data into categories. Ordinal scales also allow ranking of the categories. The most common are interval scales adding a quantifiable difference between the ranks, e.g., the Celsius temperature scale. Ratio scales are interval scales which furthermore include a well-defined zero point. They are the most complex. For scaling experiments, the following techniques are the most important:

- **Rank order:** The observer has to sort a bunch of samples with respect to the magnitude of the investigated attribute.

- **Graphical rating:** In this case, the observer is asked to rate each sample directly on an interval scale with defined end points.

- **Category scaling:** The designer of the experiment generates various categories which should not be too obviously dedicated to the samples. The observer distributes the samples into these categories. A lot of observers and stimuli are required for good statistics.

- **Paired comparisons:** Always two samples are shown to the observer who has to decide which sample is greater in the desired attribute. Tie decisions – if the observer has no preference – may also be allowed. Usually – if the number of samples is small – all possible combinations of pairs are judged to derive scales. But even if not all possible combinations are judged, quality statements or scales are possible.

To design experiments for image quality, the presence of a reference image as well as the temporal presentation of the images (simultaneously or successive) should also be considered. In this thesis, subjective data involved in the investigations are mainly based on paired-comparison experiments.

4.2 Analysis of Subjective Data

Each experimental method needs a suitable analysis of the subjective data to get the desired information. The IQA databases introduced later in this chapter provide subjective data obtained solely by scaling experiments. Hence, I present in this section two concepts of how to analyze the results of scaling experiments with respect to the performance of *image-difference metrics* (IDMs, see Chapter 3): *mean opinion scores* and *hit rates* with an additional significance analysis.

4.2.1 Mean Opinion Scores

The standard method to provide subjective scales for images of an IQA database based on scaling experiments is the calculation of so-called *mean opinion scores* (MOS). The MOS is an estimate of the actual image quality and is assigned to each image of the database. The determination of MOS – in general an average of the subjects' evaluations – depends on the experimental design. Examples of how to derive MOS from scaling experiments are given in [109, 129]. In many cases, the IQA databases provide MOS rather than raw data because MOS are sufficient for the common performance tests of IDMs and the user does not need to determine the MOS which, moreover, ensures that every user utilizes the same MOS.

To test the performance of an IDM on an IQA databases, the image-difference prediction of every distorted image with respect to its reference image has to be computed. If MOS are available, a correlation

between the IDM's predictions and the MOS is calculated. Usually, Spearman [135] or Kendall [63] rank-order correlation or the renowned linear Pearson correlation coefficient are utilized [108, 129]. The higher the correlation to MOS, the better is the IDM's performance and thus its correlation to human perception.

4.2.2 Hit Rates

For paired-comparison experiments, an alternative method to MOS called *hit rate* can be used instead to test the performance of IDMs on subjective data. The hit rate assesses directly the performance of an IDM on an IQA database without deriving subjective scales of the distorted images. The requirement for a hit-rate analysis is that the decisions for all paired-comparison evaluations are known.

Lissner et al. [76] described in detail the concept of hit rates as well as a corresponding significance analysis. Since I contributed to that paper as a co-author, the term *we* is used in the following to explain our findings.

We pointed out that the concept of MOS has some disadvantages:

- The calculation of MOS from paired-comparison experiments requires an assumption for the statistical model of the evaluations' distribution [38], such as the model by Thurstone [141] or Bradley-Terry [15]. But it might not be safe to assume a certain statistics.

- Furthermore, it is not straightforward to include inter- and intra-observer variability into the MOS and, again, a model has to be assumed. This information is important to determine the statistical significance between two performance results.

- Usually, on an IQA database based on paired-comparison experiments, several distortions are applied to each reference image. The resulting distorted images are compared with each other – but only those derived from the same reference. The MOS is then computed for each reference image but cannot be properly compared with the MOS which results from a different reference image. E.g., a color distortion caused by gamut mapping (see Section 2.4) on a low chromatic image is mostly less disturbing than on a high chromatic image. If the quality of a certain distorted low chromatic image is much worse compared to the others, it has a very low MOS. And if the distorted high chromatic images are of similar quality, they share a medium MOS. However, even though the worst distorted low chromatic image has a lower MOS than each of the distorted high chromatic images, it still may be perceptually closer to the reference image.

- Here, I add one of my findings which was not mentioned in the paper: It is required for the calculation of MOS from paired-comparison experiments that, for the same reference, each distorted image is comparable to another distorted image [90]. I.e., image $\mathbf{X}$ can be compared with image $\mathbf{Y}$ either by direct paired comparison ($\mathbf{X}$ vs. $\mathbf{Y}$) or by indirect comparison (e.g., $\mathbf{X}$ vs. $\mathbf{Z}$ and then

Z vs. **Y**). Hence, if some images of the same reference are not comparable like in [159], another methodology of assessing the IDM's performance is needed to exploit all subjective evaluations.

That is why we introduced hit rates for testing the prediction performance of an IDM on IQA databases based on paired-comparison experiments. The hit rate $\hat{p}$ is defined as the ratio of correctly predicted evaluation results $\hat{r}$ to the number of all evaluations $\hat{n}$:

$$\hat{p} = \frac{\hat{r}}{\hat{n}}. \tag{4.1}$$

Correctly predicted means that the IDM predicts a smaller difference to the reference image for the image of the paired comparison that was also rated better by the subject. We omitted tie decisions because it is almost impossible for IDMs to predict ties.

Since the concept of hit rates is not very common in IQA, we presented some general properties:

- The worst possible prediction performance is indicated by a hit rate $\hat{p} = 0.5$. This should be the case for a totally random predictor. For values lower than 0.5 the prediction for each pair can be inverted to get values higher than 0.5.

- The best possible prediction performance, i.e., the maximal achievable hit rate on a database, is called *majority hit rate* $\hat{p}_m$. Generally, the most intuitive value is not $\hat{p}_m = 1$. Due to inter- and intra-observer variability, the choices for an image pair evaluated by several observers may not be unambiguous. The majority hit rate is achieved by correctly predicting the image that the majority preferred for each image pair.

- The hit rate of an IDM correlates linearly to the Kendall correlation [63] if each image pair is evaluated exactly once. In this case, the range of hit rates is from 0 to 1 whereas the range of Kendall correlations is from -1 to 1.

For calculating the hit rate of an IDM, the knowledge of all evaluations is necessary which has to be provided by the database. For the described databases in the subsequent sections, this is only the case for the gamut-mapping databases. Nonetheless, the advantage of hit rates is that no population model has to be assumed to derive any characteristic values for the images because solely the raw experimental data is needed. Another big advantage is the determination of significance for a certain results as described in the end of this section.

4.2.3 Hit-Rate Ratio

For an easier valuation of a hit rate on a database, a linear transformation of the hit rate may be beneficial. We proposed for that purpose the ratio $\hat{p}/\hat{p}_m$ which is not affected by inter-observer and intra-observer

variability. However, this ratio cannot be compared on different databases. E.g., a good hit rate $\hat{p} = 0.7$ on a database with majority hit rate $\hat{p}_m = 0.75$ leads to the same ratio of 0.93 as a bad hit rate $\hat{p} = 0.51$ on a database with $\hat{p}_m = 0.55$.

Instead, I propose in this thesis my approach of the *hit-rate ratio* $\hat{p}_r$:

$$\hat{p}_r = \frac{\hat{p} - 0.5}{\hat{p}_m - 0.5} \tag{4.2}$$

with values between -1 and 1. This ratio is similar to correlations because a ratio of 0 means the worst possible prediction performance and a value of -1 or 1 the best possible prediction performance. Thus, my proposed hit-rate ratio $\hat{p}_r$ is easier to evaluate and allows a hit-rate comparison on different databases. In the given example, the good hit rate has a hit-rate ratio $\hat{p}_r = 0.8$ while the bad hit rate has a hit-rate ratio $\hat{p}_r = 0.2$.

4.2.4 Significance Analysis of Hit Rates

For the comparison of two IDMs, it is not sufficient to only determine their hit rates. More conclusive is the statement that the hit-rate difference of two IDMs is *significant*. Otherwise, a hit-rate difference may have happened by chance. In the following, a significant analysis of hit rates proposed in [76] – which I co-authored – is introduced.

We assumed a binomial distribution for the IDMs' predictions of observer decisions. In this way, the prediction of a single decision corresponds to a success/failure experiment – called Bernoulli experiment – with unknown success probability p_1 and p_2, respectively. I.e., each observer decision is correctly predicted with probability p_1 by the first IDM and with p_2 by the second IDM. The respective estimates of the success probabilities are denoted as $\hat{p}_1$ and $\hat{p}_2$ which correspond to the hit rates.

Suppose the first IDM correctly predicts $\hat{r}_1$ decisions out of $\hat{n}_1$ decisions and the second IDM correctly predicts $\hat{r}_2$ out of $\hat{n}_2$. Then, according to Yule's two sample binomial test [17], the confidence interval CI for $p_1 - p_2$ is [76]:

$$CI \quad = \quad [\hat{p}_1 - \hat{p}_2 - \Psi, \hat{p}_1 - \hat{p}_2 + \Psi] \tag{4.3}$$

with

$$\hat{p}_1 = \frac{\hat{r}_1}{\hat{n}_1} \tag{4.4}$$

$$\hat{p}_2 = \frac{\hat{r}_2}{\hat{n}_2} \tag{4.5}$$

$$\Psi = z_{\gamma/2}\sqrt{\left(\frac{1}{\hat{n}_1}+\frac{1}{\hat{n}_2}\right)\bar{p}\bar{q}} \tag{4.6}$$

$$\bar{p} = \frac{\hat{r}_1+\hat{r}_2}{\hat{n}_1+\hat{n}_2} \tag{4.7}$$

$$\bar{q} = 1-p \tag{4.8}$$

$$z_{\gamma/2} = \left\{s \ : \ \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s} \mathrm{e}^{-t^2/2}\mathrm{d}t = 1-\gamma/2\right\} \tag{4.9}$$

where γ is the significance level, and $z_{\gamma/2}$ is the upper $\gamma/2$-th quantile of the standard normal distribution (usually to look up in the standard normal table). In most applications, the hit rates are computed on the same set of images, i.e., $\hat{n}_1 = \hat{n}_2 = \hat{n}$. Finally, the hit rates $\hat{p}_1$ and $\hat{p}_2$ are proved to be significantly different if $0 \notin CI$. In this thesis, a significance level $\gamma = 0.05$ is assumed for the significance analysis of hit rates. This is a commonly used significance level in statistics [124]. Other frequently used levels are $\gamma = 0.1$ or $\gamma = 0.01$ – depending on the application. The lower the significance level, the more significant is the result and the harder is to fulfill the condition of significance.

With the knowledge of how to analyze subjective data, I present in the subsequent sections IQA databases which employ these methods. The IQA databases are divided into two groups depending on the type of distortion used to create the databases' images: conventional distortions and gamut-mapping distortions.

4.3 Conventional Distortions

In digital imaging, a color image is usually modified at many stages, especially:

- acquisition, e.g., noise, blur, lens corrections,
- image processing, e.g., image enhancement, denoising,
- image storage, e.g., JPEG compression,
- image transmission, e.g., transmission and quantization errors,
- displaying, e.g., gamma correction.

I refer to these commonly occurring image distortions as *conventional distortions*. In the following, I present some publicly available IQA databases with conventional distortions which are renowned in the image quality community.

4.3.1 LIVE Image Quality Assessment Database

The *Laboratory for Image and Video Engineering* (LIVE) at *The University of Texas at Austin* published an extensive IQA database in 2006 [129] called *LIVE database*. The database is publicly available online [130]. It includes 29 high-quality natural reference color images with a diversity of scenes as shown in Figure 4.1.

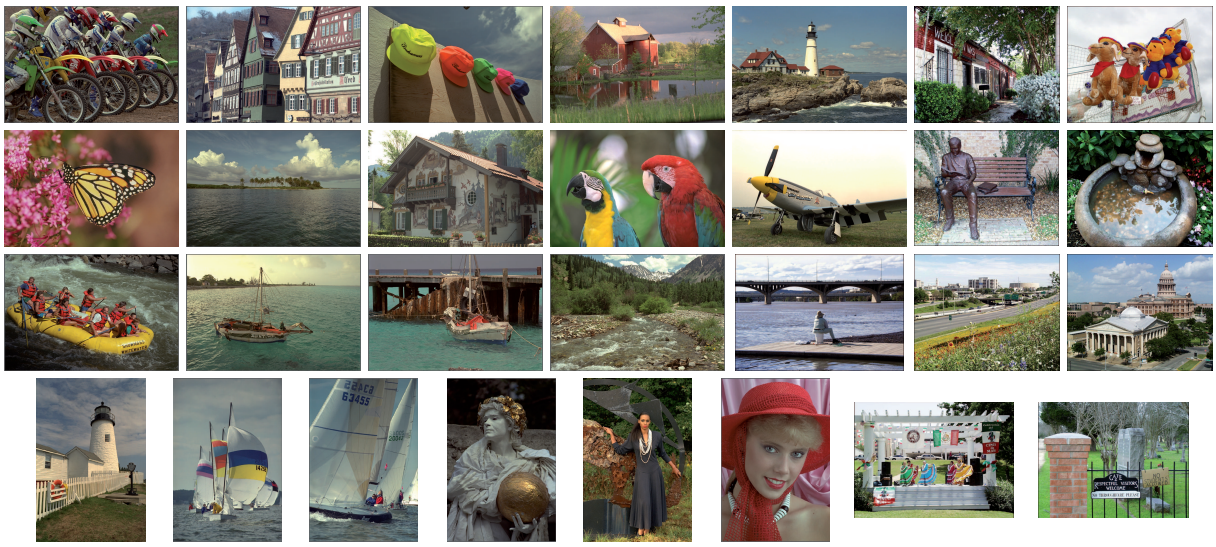


Figure 4.1: The 29 reference images of the LIVE database [129].

Five different distortion types with different strengths were used to build up the database, namely JPEG compression, JPEG2000 compression, white noise, Gaussian blur, and fast fading (bit errors during simulated wireless transmission). The image modifications result in 779 distorted images.

The methodology of the experiment was a single-stimulus quality scaling, i.e., only one image (reference or distorted) was shown and the subject had to rank the quality with a slider between ”bad” and ”excellent”. The experiment took place in an office environment, the images were shown on 21-inch CRT monitors with a resolution of $1\,024 \times 768$ pixels. The average number of subjects per ranked image was about 23. In total, more than 25 000 human image quality evaluations were collected.

After the experiment, the raw data of all evaluations was processed in order to obtain the MOS[1]. The determination of MOS for the LIVE database is explained in detail in [129] including, e.g., outlier detection and subject rejection algorithm. Please note that in the LIVE database the MOS is a relative quantity rather than an absolute quantity as the scores of the distorted images were subtracted from the scores of the reference images.

[1]In case of the LIVE database they are called DMOS referring to difference scores because the reference image is also evaluated. Since DMOS is a sort of mean opinion score I use the more general MOS.

4.3.2 Tampere Image Database

Up to now, the world's largest publicly available IQA database was developed by Ponomarenko et al. at the *Tampere University of Technology*, Finland. The database consists of 25 reference images – 24 natural images from the Kodak Lossless True Color Image Suite [45] and one artificial image – shown in Figure 4.2. The images were modified so that they all have the same resolution. In 2008, a first version was released [109] and is called *Tampere Image Database 2008* (TID2008). An extended version was published in 2013 [107, 108] – thus named *Tampere Image Database 2013* (TID2013).

Figure 4.2: The 25 reference images of both TID databases [107, 109].

The TID2008 contains 17 distortion types with four levels of distortion whereas TID2013 contains seven additional distortion types and five levels of distortion. This leads to a total number of 1 700 distorted images for TID2008 and 3 000 distorted images for TID2013, respectively. The distortion types are summarized in Table 4.1. The Tampere Image Database includes not only common distortions like compression or noise but also exotic distortions like local block-wise distortions of different intensity.

The subjects were asked to evaluate the distorted images in paired-comparison experiments also showing the reference image. The experiments took place at different locations and on the Internet so that the viewing conditions are not consistent and well defined. Due to the large number of distorted images, it is not feasible to compare each distorted image with another. Instead, the Swiss chess tournament system [109] was utilized. I.e., every turn each distorted image of a reference image is evaluated once in a paired comparison (number of paired comparisons per turn is therefore half the number of the used distorted images). The winner of a comparison gets one point, the other zero points. Nine turns are realized leading to an integer score between 0 and 9 for each distorted image at the end of the experiment. As a constraint, no paired comparison appears more than once and only distorted images with the same

Table 4.1: Summary of the Distortion Types in the Tampere Image Database of the 2008 Version [109] and the 2013 Version [107]. I Marked Employed Distortions by an 'x'.

Number	Distortion Type	TID2008	TID2013
1	Additive Gaussian noise	x	x
2	Additive noise in color components is more intensive than additive noise in the luminance component	x	x
3	Spatially correlated noise	x	x
4	Masked noise	x	x
5	High frequency noise	x	x
6	Impulse noise	x	x
7	Quantization noise	x	x
8	Gaussian blur	x	x
9	Image denoising	x	x
10	JPEG compression	x	x
11	JPEG2000 compression	x	x
12	JPEG transmission errors	x	x
13	JPEG2000 transmission errors	x	x
14	Non eccentricity pattern noise	x	x
15	Local block-wise distortions of different intensity	x	x
16	Mean shift (intensity shift)	x	x
17	Contrast change	x	x
18	Change of color saturation		x
19	Multiplicative Gaussian noise		x
20	Comfort noise		x
21	Lossy compression of noisy images		x
22	Image color quantization with dither		x
23	Chromatic aberrations		x
24	Sparse sampling and reconstruction		x

Table 4.2: Overview Conventional Image-Quality-Assessment Databases.

Database	Reference Images	Distorted Images	Subjects	Evaluations	Distortion Types	Distortion Levels
A57	3	54	7	–	6	3
CSIQ	30	866	35	5 000	6	$\overline{5}$
IVC	10	235	15	–	4	5
LIVE	29	779	$\overline{23}$	$\approx$ 25 000	5	$\overline{5}$
MICT	14	168	16	3 136	2	6
TID2008	25	1 700	838	256 428*	17	4
TID2013	25	3 000	971	524 340*	24	5
VCL@FER	23	552	118	11 307	4	6
WIQ	7	80	60	2 400	1	–

* : paired-comparison evaluations, the number of relative quality evaluations is twice that number
$\overline{\cdot}$: average number over all distorted images
– : actual number not available

(if not possible similar) number of points are paired. Finally, the MOS of a distorted image is the average of the scores over all observers.

In addition to the huge number of distorted images, the Tampere Image Database benefits from its vast number of observers and evaluations. The TID2008 experiment counted 838 subjects, for the TID2013 experiment the number is even higher with 971 in total. Every subject assessed the distorted images of one reference image resulting in 256 428 (838 subjects $\times$ 9 turns $\times$ 17 distortion types $\times$ 4 distortion levels / 2 for the image pairs) and 524 340 (971 $\times$ 9 $\times$ 24 $\times$ 5 / 2) paired-comparison evaluations, respectively.

4.3.3 Further Conventional Image Databases

During my research, I investigated also the following conventional IQA databases: the A57 [20], the CSIQ [68], the IVC [69], the VLC@FER [154], the WIQ [37], and the MICT Image Quality Evaluation Database [96]. Since they are not used in this work they are given only for supplement without further description. An overview of conventional image databases is given in Table 4.2.

4.4 Gamut-Mapping Distortions

In most of the conventional IQA databases, color distortions do not occur or they are only small compared to the achromatic distortions. Furthermore, many IDMs disregard color and only work on grayscale images (see Section 3.1.1). But color distortions can have a high impact on image quality especially when memory colors come into play, e.g., brand colors, skin tones [139]. In gamut mapping mostly color

distortions appear but also loss of structure in case of clipping. A further artifact of gamut mapping is lightness inversion.

Contrary to conventional distortions which operate locally in most cases, gamut-mapping distortions are mainly global distortions. Hence, conventional and gamut-mapping distortion are usually considered separately. Thus, I introduce a second type of IQA database called *gamut-mapping database*. Several *gamut-mapping algorithms* (GMAs) with respect to one or more gamuts are applied to a variety of reference images to generate the distorted images. In most cases, the gamuts are small newspaper gamuts to have visible differences in the distorted images for the visual experiment.

There do not exist many gamut-mapping databases – a small survey was given by Morovič in the book *Color Gamut Mapping* on Page 252 [88]. In the following, the Empa as well as the Dugay Gamut-Mapping Databases which are employed in this thesis are introduced.

4.4.1 Empa Gamut-Mapping Database

The largest publicly available gamut-mapping database was collected at the *Swiss Federal Laboratories for Materials Science and Technology* (Eidgenössische Materialprüfungs- und Forschungsanstalt, Empa). At the official web page [34], the database can be downloaded free of charge. It is divided into several smaller datasets which were collected from different experiments [121].

All experiments were conducted as paired comparisons and the result of each evaluation is recorded and provided by the *Empa Gamut-Mapping Database*. The knowledge of all evaluations has the advantage that the hit rate instead of MOS can be used to test the performance of IDMs. The determination of subjective scores for the distorted images was therefore left out. To facilitate the hit-rate approach, tie decisions are omitted on all datasets. A short description of all datasets within the Empa Gamut-Mapping Database and their release year is given in the following:

- **Algorithm Mixing (2011):** Zolliker et al. [157] introduced image fusion in order to optimize gamut mapping. Several GMAs are applied to a reference image to get a variety of gamut-mapped images. The reference image is segmented into patches and for each patch the best gamut-mapped image with respect to an image-difference metric is chosen. The fused in-gamut images show a higher perceived quality than the images obtained from gamut mapping only. The psychophysical experiment to verify the approach concludes six datasets, called *AlgMix 1* – *AlgMix 6*. Five to eight GMAs were applied to 36 and 50 reference images, respectively.

- **Conjoint Analysis (2010):** A study of parametrized gamut mapping with up to five parameters was accomplished by conjoint analysis [159], a method used in market research. It is shown that an individualized gamut mapping is possible if enough data is available. Such an approach generates an enormous amount of distorted images which was evaluated in a laboratory and a web environment.

Two datasets were built from 85 and 95 reference images and are denoted as *Conjoint Analysis 1* and *Conjoint Analysis 2*, respectively.

- **Basic Study (2010):** A comparison of 2 state-of-the-art GMAs (HPminDE, SGCK [25]) with 9 image-dependent and spatial GMAs is presented by Barańczuk et al. [8]. The dataset is a traditional benchmark study and thus named *Basic Study*. Over 5 000 evaluations were conducted for 1 067 color distorted images based on 97 reference images.

- **Image Gamut (2007):** Every image possesses a specific image gamut defined by all its colors. Mapping of the image gamut instead of the more general device gamut is supposed to have higher quality mapping results. This concept of image-dependent gamut mapping was validated by Giesen et al. in a psychophysical experiment [49]. The study called *Image Gamut* consists of 65 reference images, 520 distorted images, 8 different GMAs, and 3 698 evaluations made by 42 subjects.

- **Local Contrast (2007):** The approach of enhancing local contrasts of already gamut-mapped images was introduced by Zolliker and Simon in 2007 [158]. The good performance was verified by comparing 4 global GMAs with and without contrast enhancements. 72 reference images leading to 576 distorted images were used in the experiment. A total of 5 209 evaluation were made by 21 subjects.

4.4.2 Dugay Gamut-Mapping Database

At the Norwegian Color Research Laboratory in Gjøvik, the accuracy of five GMAs was evaluated by Dugay et al. [31] in 2008. Some aspects of perceived image quality were investigated: expert vs. non-expert opinion, display vs. hard copy, image scene dependency. For the psychophysical experiment, the five GMAs were compared with each other as paired comparisons. 20 reference images were chosen, hence 100 distorted images are included in the database called *Dugay Database*. The database is reliable because 20 observers evaluated each pair twice (with interchanged distorted images) in random order leading to 8 000 paired-comparison evaluations.

4.5 Combined Gamut-Mapping Database

In this section, I propose a combination of the presented gamut-mapping databases (see Section 4.4) to get a large database for investigations of gamut-mapping distortions.

For training and testing of image-difference metrics, a large image-quality-assessment database with good properties (many scenes, sufficient statistics) is beneficial. In the case of gamut-mapping distortions, the combination of the Empa and the Dugay Database offers such a large database. However, not every dataset of the Empa Database has highly reliable experimental results and should be neglected. Especially

web-based experiments have a high fraction of outliers, e.g., some subjects always clicked left, and the statistical properties of the Conjoint Analysis 1 and 2 lack because many distorted images are only evaluated once.

Thus, I propose the *Combined Gamut-Mapping Database* which is the combination of the Dugay Database (see Section 4.4.2) and AlgMix 1, AlgMix 2, Basic Study, Image Gamut, and Local Contrast from the Empa Database (see Section 4.4.1). It consists of 277 reference images, from 5 up to 19 GMAs applied leading to 2 479 distorted images, and 29 665 subjective evaluations. As the Combined Gamut-Mapping Database is a combination of different sets, duplicate images were removed while keeping the subjective evaluations. The actual number of observers cannot be counted because it is not known if some observers contributed to several datasets. I estimate a number of 125 observers by assuming 10 observers for Basic Study and that the observers of AlgMix 1 and Local Contrast are included in the number of observers of AlgMix 2 and Image Gamut, respectively.

Please note that in general such combinations are not possible for the databases of conventional distortions. The acquisition of MOS is different for each database and therefore not comparable to other databases. In contradiction, the gamut-mapping datasets include the raw data of every paired-comparison evaluation. Hence, the computation of hit rates $\hat{p}$ is still valid for the combination of gamut-mapping databases. However, if equal images originating from different datasets are compared, the majority hit rate may be overrated and should be recomputed.

All gamut-mapping datasets are summarized in Table 4.3 also including the Combined Gamut-Mapping Database. In the subsequent chapters, only the Combined Gamut-Mapping Database is used for the evaluation of gamut-mapping distortions. Some parts of the next chapter describe how a selection of IDMs performs on the Combined Gamut-Mapping Database and how this database helps to determine parameters of an IDM.

Table 4.3: Overview Gamut-Mapping Image-Quality-Assessment Databases and Composition of the Combined Gamut-Mapping Database.

Dataset	Reference Images	Distorted Images	Subjects	Evaluations*	GMAs
AlgMix 1 [157]	36	216	11	3 900	6
AlgMix 2 [157]	36	180	53	3 659	5
AlgMix 3 [157]	36	216	12	4 713	6
AlgMix 4 [157]	50	400	9	4 869	8
AlgMix 5 [157]	36	216	99	3 739	6
AlgMix 6 [157]	50	400	149	5 123	8
Conjoint Analysis 1 [159]	85	5 320	701 once, 41 twice	2 860	–
Conjoint Analysis 2 [159]	95	12 986	120	11 401	–
Basic Study [8]	97	1 067	9 - 12 (–)	5 199	11
Image Gamut [49]	65	520	42	3 698	8
Local Contrast [158]	72	576	21	5 209	8
Dugay [31]	20	100	20	8 000	5
Combined	277	2 479	125 (–)	29 665	5 - 19

*: paired-comparison experiments, tie decisions were omitted

–: actual number not available

datasets in bold face are included in the Combined Gamut-Mapping Database

5

The Color-Image-Difference (CID) Metric

In this chapter a new image-difference metric is proposed – the *Color-Image-Difference* (CID) metric. It has a modular framework which allows to take into account viewing conditions, hypotheses on the human visual system, color information, and image-difference features. In the following, I explain why a new image-difference metric is needed and how the modular framework is built and its parameters are trained. The new metric is tested on gamut-mapping distortions leading to the best result among the investigated metrics. On conventional distortions, it shows a high performance as well.

5.1 Related Publications

The content of the following chapter is mainly based on the journal paper by Lissner et al. [76]:

> I. Lissner, J. Preiss, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. Image-Difference Prediction: From Grayscale to Color. *IEEE Transactions on Image Processing*, 22(2):435–446, 2013.

It is an extension of two previously published conference papers, [113] and [75]:

> J. Preiss, I. Lissner, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. The Impact of Image-Difference Features on Perceived Image Differences. In *CGIV 2012 – 6th European Conference on Colour in Graphics, Imaging, and Vision*, pp. 43–48, 2012

and

> I. Lissner, J. Preiss, and P. Urban. Predicting Image Differences Based on Image-Difference Features. In *IS&T/SID, 19th Color and Imaging Conference*, pp. 23–28, 2011.

Most investigations are taken from these papers and are introduced in this chapter. Reference [76] as the most recent and most profound paper should thus be cited for the proposed image-difference metric. Further investigations, results, and discussions done by me for this thesis are pointed out in particular.

As co-author, I have been contributing essentially to these papers in close cooperation with other authors. To make a clear distinction between my contributions and those of the other co-authors, I use the term *I* for my work and the term *we* whenever other authors were involved. At the end of the chapter my main contributions are summarized. Please take into account that I am rewriting the previously published papers and may use the same phrasing to avoid diminishing the meaning – this is common practice as shown in Lissner's thesis [74].

5.2 The Need of a New Image-Difference Metric

There are still open questions in *image-quality-assessment* (IQA) research [19]. Among those questions I would like to address two problems in particular: color and viewing conditions. The details of those problems and an approach to solve them are given in this section.

5.2.1 Color in Image-Difference Metrics

Existing surveys of IQAs (see Chapter 3) show that far more than 100 IQAs have been proposed so far [5, 19, 101, 147]. Furthermore, most of them are full-reference IQAs denoted as *image-difference metrics* (IDMs) and the majority disregard color information (see Section 3.2). The omission of color has the advantage of fast computation and less storage of data in IQA applications. Surprisingly, the prediction performance of selected grayscale IQAs is remarkably high on most distortions used in image processing such as blur, noise, or compression artifacts [129].

However, there are also distortions that cannot be detected by grayscale IQAs. These are, in particular, distortions which keep the lightness channel unchanged like gamut-mapping distortions or change of saturation. An example of such a distortion is shown in Figure 5.1. An IDM would be beneficial which is sensitive to chromatic distortions and furthermore shows a high accuracy in predicting the achromatic conventional distortions. Thus, I suggest to derive a color-image-difference metric from an existing and well performing grayscale IDM.

As color is not predominantly considered in IQA research, investigations of color-distortion databases are rarely addressed in literature. In addition to the creation of such databases, there are also missing performance tests of a variety of IDMs on color-distortion databases. Such a performance test will be carried out later in this chapter.

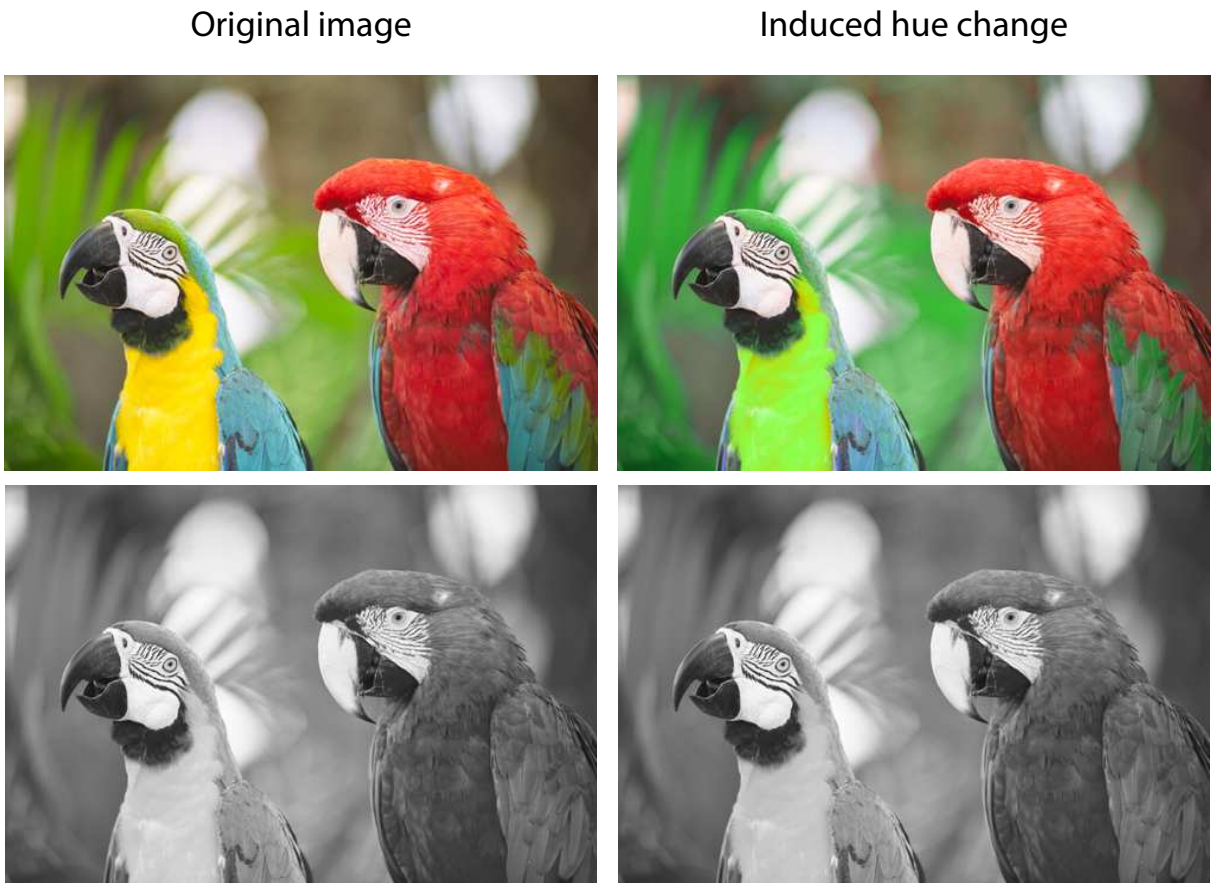


Figure 5.1: Two color images only differing in a chromatic distortion induced by a hue change (first row) and their grayscale versions (second row). Even though the distortion is clearly visible on the color images, it is not so on the grayscale images which would lead to a misinterpretation of a grayscale IDM. The original image is taken from the Kodak Lossless True Color Image Suite [45].

5.2.2 Viewing Conditions in Image-Difference Metrics

Images are viewed under various conditions in everyday life, for instance, in the office in front of the monitor, at the cinema, or outside on a mobile device. Viewing conditions include, in particular, visual resolution, surrounding illumination, and luminance level as well as viewing angle, illuminant, and so on. For the evaluation of image quality, the viewing conditions have to be known as in the example of Figure 5.2. However, most metrics disregard viewing conditions of the input images. They just take two images as an input independent of their viewing conditions. Thus, those IDMs need to assume standard viewing conditions for the images. Normally, an office environment is assumed, i.e., indoor illumination, visual resolution of about $VR = 40$ samples per degree of visual angle ($VR = 40\,\mathrm{s/deg}$), and a white point luminance of about $100\,\mathrm{cd/m^2}$.

Models about the influence of viewing conditions on image quality already exist – so-called *image-*

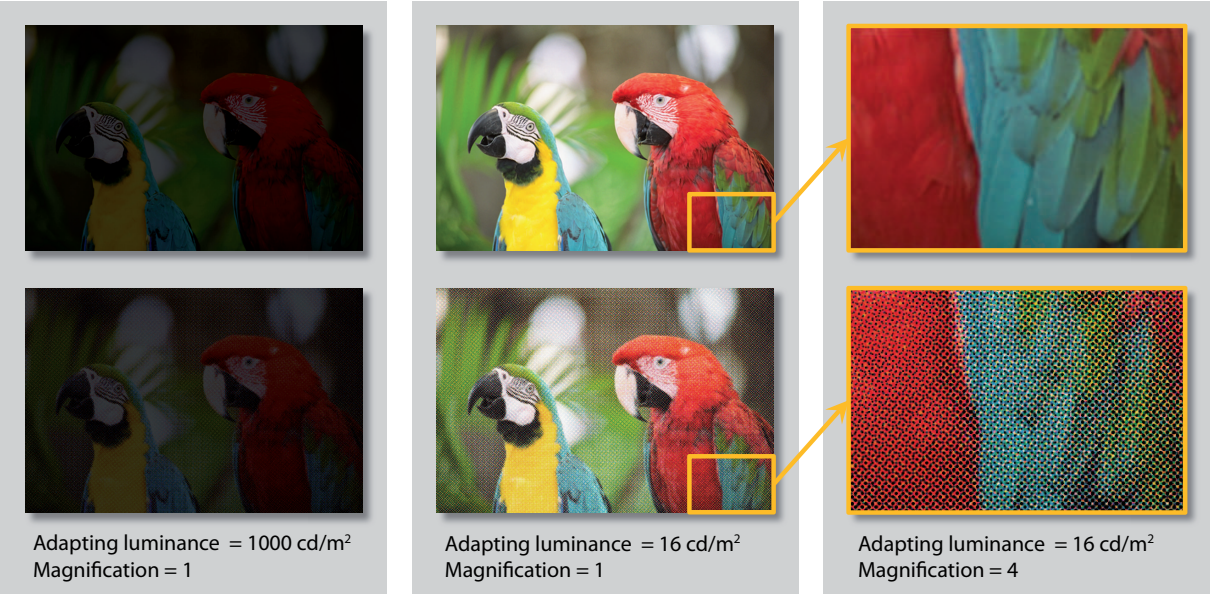


Figure 5.2: Different viewing conditions lead to a distinct perception of image differences. The first row shows continuous-tone images and the second row shows half-tone images displayed with a white point luminance of 80 $\mathrm{cd/m^2}$. Left column simulates outdoor daylight environment (adapting luminance of 1000 $\mathrm{cd/m^2}$) while the middle and the right column simulate office environment (adapting luminance of 16 $\mathrm{cd/m^2}$). The right column shows a four-fold magnification of an image part extracted from the middle column. The subthreshold image difference in the left column turns into a suprathreshold image difference in the middle column by changing the adapting luminance. The magnification in the right column simulating a closer viewing distance further increases the image difference. The figure has originally been published in [76].

appearance models (IAMs). A detailed explanation about IAMs is given in [38]. Chromatic adaptation, contrast sensitivity, appearance phenomena like the Hunt effect or the Stevens effect as well as other mechanisms describing the *human visual system* (HVS) are considered by IAMs. But IAM research is still in its infancy and only a few IAMs have been proposed – a renowned example is *iCAM06* [66].

The transformation of an image into standard viewing conditions by an IAM can be interpreted as a *normalization* step. For an accurate comparison of images under different viewing conditions a normalization step should be applied before evaluation. If desired, the images are transformed into a perceptual working color space as the last step of the normalization. The concept of normalization offers a good prospect to be utilized in IQA for considering viewing conditions.

The working color space is supposed to give an easier access to the images' color properties for the evaluation. We suggested to use an opponent color space which provides lightness, chroma, and hue predictors. Further desired attributes of such a color space are hue linearity and perceptual uniformity. Hue linearity means that chroma and hue are free from cross contamination and perceptually uniformity that Euclidean distances in the space are proportional to perceived color differences. A perfectly perceptually uniform color space does not exist [61] but there are various approximations [26, 78, 143].

Most renowned databases as introduced in Chapter 4 do not provide detailed information about viewing conditions. At least, important specifications such as visual resolution or viewing environment are often given. The majority of the databases' experiments were conducted under general conditions – in an office environment. Even if the visual resolution is not available, a rough estimate of viewing distance and pixel pitch should be sufficient to determine the visual resolution because liberal deviations do not drastically change perception. For instance, sitting 50 cm or 75 cm in front of a display with a normal pixel pitch does not lead to a very different image perception.

To summarize, while most IDMs only have two grayscale images as input we suggested to evaluate color images and the viewing conditions. The latter are needed for normalizing the images to standard viewing conditions – office environment, average visual resolution. This normalization is usually done by IAMs. If the viewing conditions are not exactly known, a default parameter set has to be assumed or the IAM has to be restricted to a simpler model.

5.3 Image-Difference Features

To account for the open questions of color and viewing conditions described above, we presented a variable IDM framework which is more general [113]. Elaborating the framework resulted in a new IDM which we denoted as *Color-Image-Difference metric* (CID metric) [76]. After the normalization step in which an IAM is applied to color images, so-called *image-difference features* (IDFs) are extracted from the normalized images within the working color space. A general overview of the CID framework is shown in Figure 5.3.

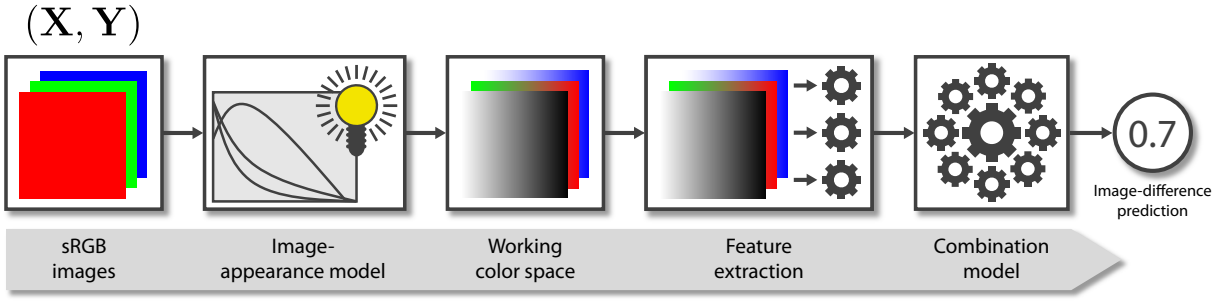


Figure 5.3: Modular framework of the CID metric. The original version of the figure has already been published in [76].

IDFs are functions which calculate characteristic values of image differences with respect to hypotheses made on the HVS (more details are provided in Section 2.1). We proposed and investigated some hypotheses [113]:

- **Hypothesis 1:** *The HVS is sensitive to lightness, chroma, and hue differences.*

  This hypothesis refers to the opponent color theory: the HVS processes an achromatic channel

(lightness) and two chromatic channels (red-green and blue-yellow corresponding to chroma and hue predictors) rather than cone responses. Techniques such as gamut-mapping algorithms (see Section 2.4) exploit this property.

- **Hypothesis 2:** *The HVS is sensitive to achromatic contrast differences.*
  Contrast masking of the visual system describes the phenomenon that spatial patterns reduce the sensitivity to detect contrast differences [19, 72]. Hence, this hypothesis is related to contrast masking.

- **Hypothesis 3:** *The HVS is sensitive to achromatic structural differences.*
  In addition to contrast masking, the effect of *pattern masking* [30, 51] – thus structural differences – plays an important role in visual processing.

- **Hypothesis 4:** *The HVS is more sensitive to contrast changes in low contrast image regions.*
  Weber's law states that the threshold difference of two stimuli is proportional to the magnitude of the stimuli. The same holds true for suprathreshold contrast differences: for instance, a contrast change causes an edge to vanish in a low-contrast region while for the same contrast change in a high-contrast region the edge is still present. However, the latter case is supposed to be less disturbing.

- **Hypothesis 5:** *The HVS is not equally sensitive to differences described in hypotheses 1–4 on different spatial frequency bands.*
  Psychophysical experiments revealed that the sensitivity to detect contrast differences depends on the spatial frequency of the visual target. *Contrast-sensitivity functions* (CSFs) are derived from these experiments for the visibility threshold level [144]. To account for the suprathreshold contrasts of complex structured images, a multi-scale approach was proposed by Wang et al. [151].

These hypotheses are rather low-level assumptions of the HVS. A more general review about visual psychophysics – the interaction of physical attributes of a visual stimuli and psychological responses – is given by Chandler [19].

The general definition of an IDF is [76]:

$$\mathbf{IDF}: \ \mathcal{I}_{M,N} \times \mathcal{I}_{M,N} \times \mathcal{P} \to [0,1] \tag{5.1}$$

where $\mathcal{I}_{M,N}$ is the set of all sRGB color images with M rows and N columns and $\mathcal{P}$ is the set of all parameter arrays describing the viewing conditions for an IAM. Therefore, an IDF is a function applied directly on input sRGB images $\mathbf{X}, \mathbf{Y}$ with respect to the viewing conditions.

For the calculation of IDFs, we proposed a modular framework which includes image normalization **N**, downscaling **D**, and the actual feature extraction **F**:

$$\mathbf{IDF} = \mathbf{F} \circ \mathbf{D} \circ \mathbf{N} \tag{5.2}$$

with

$$\mathbf{N}: \quad \mathcal{I}_{M,N} \times \mathcal{I}_{M,N} \times \mathcal{P} \quad \rightarrow \mathcal{W}_{M,N} \times \mathcal{W}_{M,N} \tag{5.3}$$

$$\mathbf{D}: \quad \mathcal{W}_{M,N} \times \mathcal{W}_{M,N} \quad \rightarrow \mathcal{W}_{\acute{M},\acute{N}} \times \mathcal{W}_{\acute{M},\acute{N}} \tag{5.4}$$

$$\mathbf{F}: \quad \mathcal{W}_{\acute{M},\acute{N}} \times \mathcal{W}_{\acute{M},\acute{N}} \quad \rightarrow [0,1] \tag{5.5}$$

where $\mathcal{W}_{M,N}$ is the set of all images with M rows and N columns represented in the working color space, $M \geq \acute{M}$, and $N \geq \acute{N}$. Image normalization into a working color space with respect to viewing conditions is described in Section 5.2.2. Due to *hypothesis 5*, scale dependency affects IDFs and is therefore included by downscaling **D**.

5.3.1 Feature Extractions

For each feature extraction **F**, we used a specific local image comparison t on corresponding $k \times k$ sliding windows ($k \ll \min(M,N)$) within the images:

$$t: \ \mathcal{W}_{k,k} \times \mathcal{W}_{k,k} \rightarrow [0,1]. \tag{5.6}$$

The image comparison t is calculated for each pair of windows $(\mathbf{x},\mathbf{y})$ of a set $\mathcal{K}$ of corresponding sliding windows within the normalized images $\hat{\mathbf{X}}, \hat{\mathbf{Y}} \in \mathcal{W}_{\acute{M},\acute{N}}$. A pair of corresponding windows $(\mathbf{x},\mathbf{y})$ are windows $\mathbf{x}$ in $\hat{\mathbf{X}}$ and $\mathbf{y}$ in $\hat{\mathbf{Y}}$ at the same pixel positions. The windows $\mathbf{x}$ and $\mathbf{y}$ are equivalent to pixel arrays containing all pixels $\mathbf{x}_i$ and $\mathbf{y}_i$, $i \in \{1,2,...,k^2\}$ within these windows. The feature **F** is then computed by averaging the local image comparisons:

$$\mathbf{F}(\hat{\mathbf{X}},\hat{\mathbf{Y}}) = \frac{1}{|\mathcal{K}|} \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{K}} t(\mathbf{x},\mathbf{y}) \tag{5.7}$$

where $|\mathcal{K}|$ is the number of windows in $\mathcal{K}$.

There are other possibilities than taking the mean to pool the values such as weighting with respect to saliency or computing a quantile. A more detailed overview on pooling strategies is given in [150]. However, we investigated only the mean which is the most common and which led to good results.

Please note that the number $|\mathcal{K}|$ of pairs of corresponding sliding windows within the compared images do not need to be the number $M \cdot N$ of image pixels. Depending on the method at the image's boundary the set $\mathcal{K}$ includes, for instance, only valid sliding windows ($|\mathcal{K}| = [M-k+1] \cdot [N-k+1]$) or valid plus virtual sliding windows ($|\mathcal{K}| > [M-k+1] \cdot [N-k+1]$). Virtual sliding windows insert virtual values outside the image's boundary for computation which is called *padding*. Padding strategies of MATLAB – e.g., mirror-reflecting the boundary pixels – are introduced in [50]. We used only valid sliding windows for the calculation of IDFs to avoid incorrect boundaries with the drawback of neglecting the boundary pixels.

5.3.2 Image-Difference Comparisons

The aim of IDFs is to extract features which highly correlate with the HVS. The renowned *Structural Similarity* (SSIM) index has a high correlation to subjective data and is made of three image comparisons (see Section 3.3.2). Hence, these image comparisons are well suited for our framework and we derived our proposed IDFs from them. The SSIM's image comparisons *luminance* l, *contrast* c, and *structure* s given in Equations 3.10 - 3.12 are computed for grayscale images.

On the contrary, we work on color images which allows to gain also chromatic information. In the working color space, an image pixel $\mathbf{x}_i$, $i \in \{1, 2, ..., k^2\}$ within a window $\mathbf{x}$ consists of lightness channel L, red-green channel a, and blue-yellow channel b: $\mathbf{x}_i = (L_{\mathbf{x}_i}, a_{\mathbf{x}_i}, b_{\mathbf{x}_i})$. The chroma predictor of the pixel is $C_{\mathbf{x}_i} = \sqrt{a_{\mathbf{x}_i}^2 + b_{\mathbf{x}_i}^2}$.

Listed below are five image-difference comparisons derived from the SSIM index [149] which lead to our proposed IDFs [76]. We chose these five image-difference comparisons which also extract chroma and hue features as suggested in [121]. Please note that we could also have used other image-difference comparisons and another number of image-difference comparisons.

- **Lightness-Difference Comparison**

  Lightness-difference comparison l_L based on *hypothesis 1*:

$$l_L(\mathbf{x}, \mathbf{y}) \quad = \quad \frac{1}{c_1 \cdot \overline{\Delta L(\mathbf{x}, \mathbf{y})}^2 + 1} \tag{5.8}$$

  where parameter $c_1 > 0$ and $\overline{\Delta L(\mathbf{x}, \mathbf{y})}$ denotes the Gaussian-weighted mean of lightness difference $\Delta L(\mathbf{x}_i, \mathbf{y}_i)$ for each pixel pair $(\mathbf{x}_i, \mathbf{y}_i)$, $i \in \{1, 2, ..., k^2\}$ in the corresponding windows. The lightness difference is defined as:

$$\Delta L(\mathbf{x}_i, \mathbf{y}_i) \quad = \quad L_{\mathbf{x}_i} - L_{\mathbf{y}_i}. \tag{5.9}$$

The comparison is derived from the SSIM's luminance l (see Equation 3.10) which, however, is designed for an intensity-linear color space. So we transformed it into the perceptually uniform working color space by assuming approximations and Fechner's law [43] as shown in Appendix A.3. Please note that the derivation assumes small color differences and that color differences in most nearly perceptually uniform color spaces correlate well to human perception only for small color differences [22]. For larger color differences – likely for gamut-mapped images – I suggest, e.g., a difference cut-off. But we assumed to deal only with small differences and left out this point of investigation.

- **Lightness-Contrast Comparison**

  Lightness-contrast comparison c_L based on *hypothesis 2*:

$$c_L(\mathbf{x},\mathbf{y}) = \frac{2\sigma_{\mathbf{x}}^L\sigma_{\mathbf{y}}^L + c_2}{{\sigma_{\mathbf{x}}^L}^2 + {\sigma_{\mathbf{y}}^L}^2 + c_2} \tag{5.10}$$

  where parameter $c_2 > 0$ and $\sigma_{\mathbf{x}}^L, \sigma_{\mathbf{y}}^L$ are the Gaussian-weighted standard deviations σ_x, σ_y (see Equation 3.8) computed on the lightness channel L.

  The comparison is adopted from the SSIM's contrast comparison c (see Equation 3.11) but computed on the lightness channel. It reflects *hypothesis 4* because contrast differences have a larger impact on low contrast regions than on high contrast regions (compare Figure A.1 in Appendix A.3 which is also valid for SSIM's contrast comparison).

  The impact of contrast masking is adjusted by parameter c_2. The contrast-difference detection threshold is small for a small parameter and increases with increasing parameter, i.e., in low-contrast regions a small parameter emphasizes contrast differences [113]. The effect is illustrated in Figure 5.4.

- **Lightness-Structure Comparison**

  Lightness-structure comparison s_L based on *hypothesis 3*:

$$s_L(\mathbf{x},\mathbf{y}) = \frac{\sigma_{\mathbf{xy}}^L + c_3}{\sigma_{\mathbf{x}}^L\sigma_{\mathbf{y}}^L + c_3} \tag{5.11}$$

  where parameter $c_3 > 0$ and $\sigma_{\mathbf{xy}}^L$ is the Gaussian-weighted covariance σ_{xy} (see Equation 3.9) computed on the lightness channel L.

  The comparison is adopted from the SSIM's structure comparison s (see Equation 3.12) but computed on the lightness channel. The covariance $\sigma_{\mathbf{xy}}^L$ allows also negative values but is mostly positive. If parameter c_3 is too small a negative lightness-structure comparison may occur. Taking

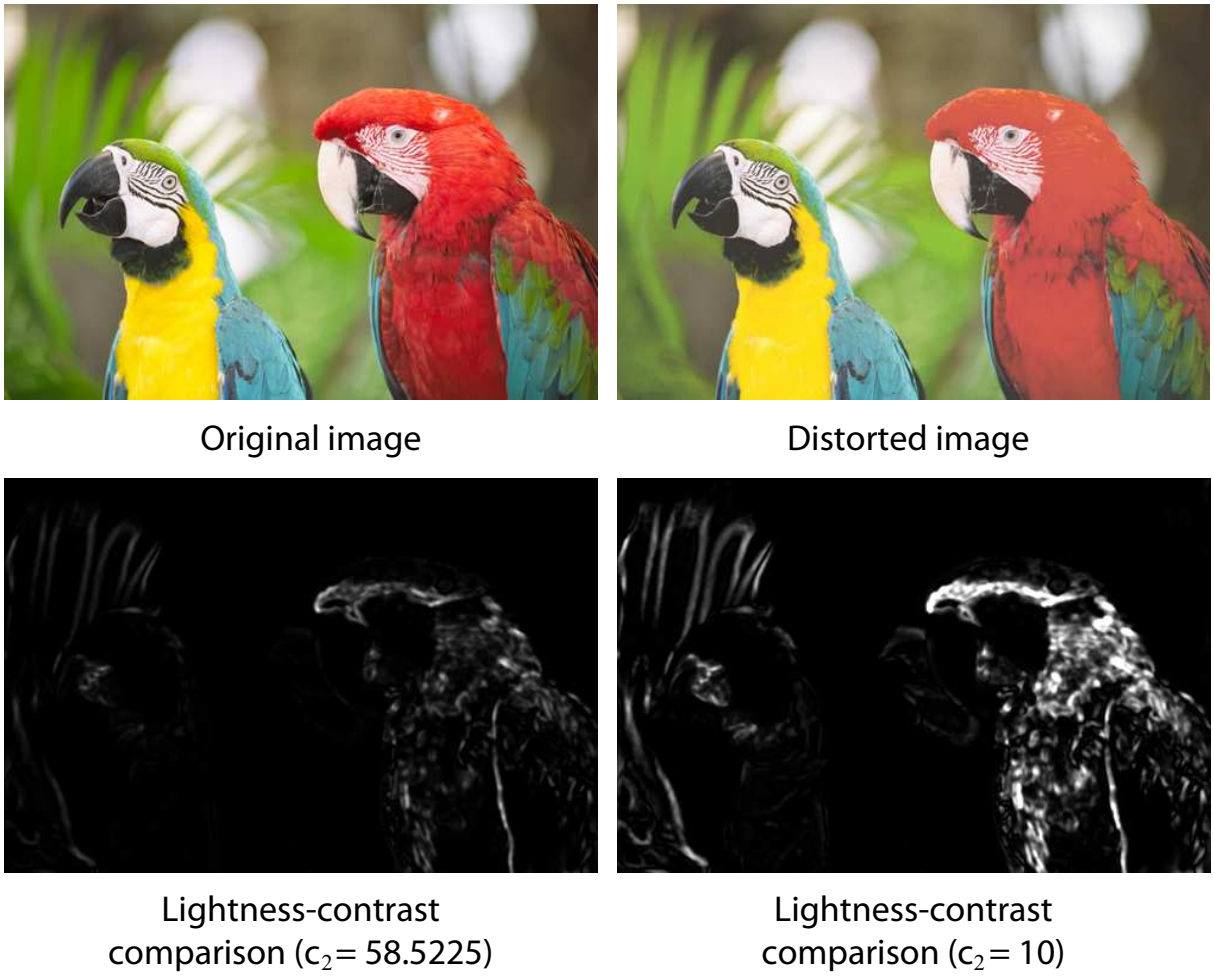


Figure 5.4: Image-difference maps (lower row) of lightness-contrast comparison l_C according to Equation 5.10 of an original and a distorted image without image normalization. Parameter c_2 adjusts the impact of contrast masking: a smaller parameter emphasizes contrast differences in low-contrast regions (e.g., red feathers). Original and distorted image are taken from the Dugay Database [31]. The original version of the figure has already been published in [76].

the absolute value $|\sigma^L_{\mathbf{xy}}|$ avoids this shortcoming but we did not consider this in the investigation due to the rare occurrence. In the SSIM's structure comparison s, the parameter c_3 is big enough to avoid negative values.

- **Chroma-Difference Comparison**

  Chroma-difference comparison l_C based on *hypothesis 1*:

$$l_C(\mathbf{x},\mathbf{y}) = \frac{1}{c_4 \cdot \overline{\Delta C(\mathbf{x},\mathbf{y})}^2 + 1} \tag{5.12}$$

where parameter $c_4 > 0$ and $\overline{\Delta C(\mathbf{x}, \mathbf{y})}$ denotes the Gaussian-weighted mean of chroma difference $\Delta C(\mathbf{x}_i, \mathbf{y}_i)$ which is defined as:

$$\Delta C(\mathbf{x}_i, \mathbf{y}_i) \quad = \quad C_{\mathbf{x}_i} - C_{\mathbf{y}_i}. \tag{5.13}$$

Contrary to the SSIM index, the proposed IDM also processes the chromatic channels – chroma and hue. Since we expected a small influence of contrast and structural differences in chromatic channels (such as noise [134]) for color applications like gamut mapping, we only regarded chroma and hue differences based on the adjusted luminance comparison in Appendix A.3.

- **Hue-Difference Comparison**

  Hue-difference comparison l_H based on *hypothesis 1*:

$$l_H(\mathbf{x}, \mathbf{y}) \quad = \quad \frac{1}{c_5 \cdot \overline{\Delta H(\mathbf{x}, \mathbf{y})}^2 + 1} \tag{5.14}$$

  where parameter $c_5 > 0$ and $\overline{\Delta H(\mathbf{x}, \mathbf{y})}$ denotes the Gaussian-weighted mean of hue difference $\Delta H(\mathbf{x}_i, \mathbf{y}_i)$ which is defined as [10]:

$$\Delta H(\mathbf{x}_i, \mathbf{y}_i) \quad = \quad \sqrt{(a_{\mathbf{x}_i} - a_{\mathbf{y}_i})^2 + (b_{\mathbf{x}_i} - b_{\mathbf{y}_i})^2 - \Delta C(\mathbf{x}_i, \mathbf{y}_i)^2}. \tag{5.15}$$

  Hue difference $\Delta H(\mathbf{x}_i, \mathbf{y}_i)$ is a Euclidean rather than a hue-angle difference because color differences are defined by their distance in a color space. For instance, for a constant hue-angle difference, the actual hue difference increases with chroma [67].

  The parameters c_1, c_4, and c_5 adjust the lightness-, chroma-, and hue-difference comparisons to the working color space. They weight lightness, chroma, and hue differences with respect to each other and have to be almost similar in a perceptually uniform color space. A larger parameter puts more weight on the corresponding difference.

With these five image-difference comparisons, we were able to create a new image-difference metric which also accounts for chromatic information. How the so-called CID metric is derived is explained in detail in the upcoming section.

5.3.3 Combining the Image-Difference Features to the Color-Image-Difference (CID) Metric

The local image-difference comparisons l_L, c_L, s_L, l_C, l_H are incorporated into individual IDFs as proposed in Equations 5.2 and 5.7. Even though image-difference comparisons and IDFs return a single value, they

are defined quite differently: image-difference comparisons operate on local windows of the normalized and downscaled images in the working color space (see Equation 5.6) whereas IDFs operate on whole sRGB images with viewing conditions as additional parameter (see Equation 5.1). To distinguish between both, I denote IDFs as $\mathbf{L}$, $\mathbf{C}$, and $\mathbf{S}$ which are based on the image-difference comparisons l, c, and s, respectively.

To account for hypothesis 5, a multi-scale approach is employed adapted from the MSSIM index (see Section 3.3.3). $m = 5$ scales are used starting with the original image size (scale 1) and low-pass filtering by a biorthogonal 9/7 wavelet and then downsampling by factor 2 to determine the next smaller scale. Contrary to MSSIM, smaller scales than scale 1 are used only for the lightness-contrast IDF $\mathbf{C}_L$ and the lightness-structure IDF $\mathbf{S}_L$.[1] The i-th scale, $i \in \{1, ..., m\}$ of the lightness-contrast and the lightness-structure IDF is denoted as ${\mathbf{C}_L}^i$ and ${\mathbf{S}_L}^i$, respectively.

In the context of the presented framework, an *image-difference metric* (IDM) is a combination of several IDFs which is supposed to have a better prediction performance than a single IDF. The structure of an IDM is the same as for an IDF (defined in Equation 5.1) – input parameters are color images $\mathbf{X}, \mathbf{Y}$ and viewing conditions described by the parameter array $\mathbf{P} \in \mathcal{P}$. The combined IDFs share the same normalization $\mathbf{N}$ (see Equation 5.3). In the following, the arguments $(\mathbf{X}, \mathbf{Y}, \mathbf{P})$ for IDFs and IDMs are omitted for the sake of clarity.

We proposed the *Color-Image-Difference metric* called CID metric by combining the presented IDFs. We employed a factorial combination model by just multiplying the IDFs and weighting the different scales:

$$\mathbf{CID} = 1 \; - \; \mathbf{L}_L \cdot \prod_{i=1}^{m} \left({\mathbf{C}_L}^i \cdot {\mathbf{S}_L}^i \right)^{\beta_i} \cdot \mathbf{L}_C \cdot \mathbf{L}_H \tag{5.16}$$

where $\mathbf{L}_L$ is the lightness-difference, $\mathbf{C}_L$ is the lightness-contrast, $\mathbf{S}_L$ is the lightness-structure, $\mathbf{L}_C$ is the chroma-difference, and $\mathbf{L}_H$ is the hue-difference IDF. The scale weights $\beta_i, i \in \{1, ..., m\}$ adopted from the MSSIM index [151] are given in Table 3.2. Please note that they are based on psychophysical studies assuming a visual resolution of $VR = 32\,\mathrm{s/deg}$ (details in Section 3.3.3). Further fixed parameters are the side length $k = 11$ of the $k \times k$ sliding windows on which the image-difference comparisons are computed and the standard deviation $\sigma_G = 1.5$ of the circular-symmetric Gaussian weighting w used for averaging in the image-difference comparisons. Both parameters are adopted from SSIM [149] from which the CID metric is derived. The remaining parameters not yet determined are the parameters c_1,...,c_5 which, in the following, I will refer to as the CID parameters.

[1] Please note that in the journal paper [76] it is written that the lightness-difference IDF $\mathbf{L}_L$ is used on the smallest scale $m = 5$. That is only correct for the official code in supplementary material. In the computations for the journal paper, $\mathbf{L}_L$ is only used on scale 1. However, an analysis revealed that the difference in the final results is not significant.

The subtraction of the combined IDFs from 1 affirms the meaning of a difference, i.e., the CID metric returns 0 if there is no perceptual difference between the images $\mathbf{X}$ and $\mathbf{Y}$. We also investigated an additive and a hybrid (mixing additive and factorial) combination model [113] as well as a polynomial model [75]. However, no significant difference between all those models was found so we decided to keep the factorial combination model from MSSIM. As an overview, the structure of the CID metric is summarized in Figure 5.5.

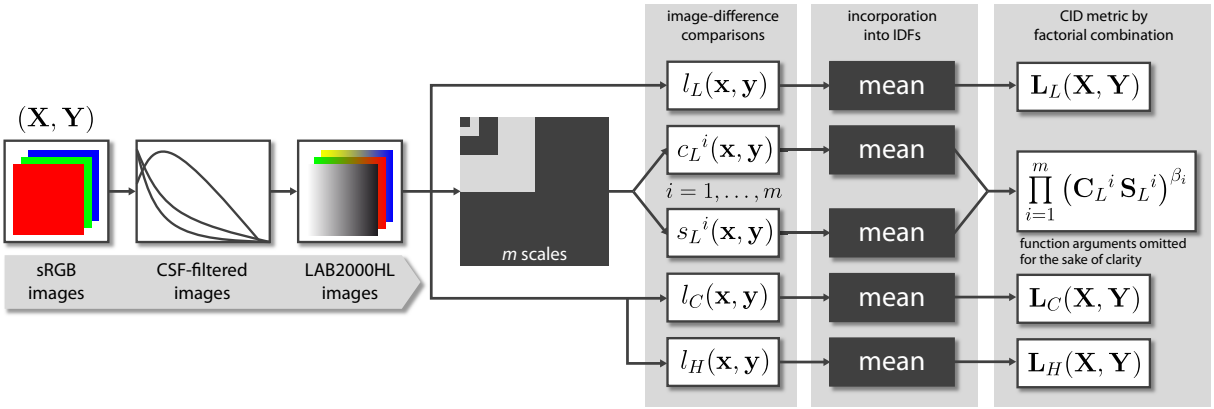


Figure 5.5: Structure of the CID metric based on individual image-difference features (IDFs). The specific image-appearance modeling (CSF filtering) and working color space (LAB2000HL) are introduced in Section 5.4. The original version of the figure has already been published in [76].

The CID parameters $c_1, ..., c_5$ are needed to adjust the CID metric to the working color space, to weight lightness, chroma, and hue differences with respect to their importance, and to regulate the influence of contrast masking. How to determine the CID parameters is described in the next section.

5.4 Elaborating the CID Metric

This section first describes which working color space and which image-appearance model are used for CID. Then, the CID parameters are elaborated. We used the common methodology of parameter training on subjective data to determine the CID parameters. Essential for parameter training is an *image-quality-assessment database* (IQA database, see Chapter 4) which includes distorted images with subjective scales. The parameters are trained on the subjective data such that the CID metric shows the best prediction performance on the IQA database. We used the concept of *hit rates* (see Section 4.2.2) to assess the subjective data of the employed IQA database.

5.4.1 Working Color Space

The requirement of the working color space is ideally an opponent color space which is perceptually uniform and hue linear. A color space designed to approximately possess these properties is the LAB2000HL

color space [78] which is therefore used for the CID metric. Its perceptual uniformity is induced by the CIEDE2000 color-difference formula [23] and hue linearity by the Hung and Berns data of constant perceived hue [53].

Further color spaces which partly fulfill the desired requirements are the hue linear *IPT* color space [32] or the color space used in *CIECAM02* [24]. However, they are not considered for the CID metric as only LAB2000HL covers both perceptual uniformity and hue linearity.

5.4.2 Image-Appearance Model

Since most information about viewing conditions is not provided as metadata by the databases, we limited the normalization to simple IAMs which only consider the visual resolution. The alternative of assuming a default parameter set was not considered as other IAMs are also not so reliable. For luminance level and illuminant, an office environment is assumed just like the LAB2000HL color space was designed for. Hence, the visual resolution was assumed to have the most important impact. The impact of visual resolution is modeled by filtering images with the *contrast-sensitivity functions* (CSFs) of the HVS which correspond to the inverse contrast-detection threshold. A more detail description of CSFs is presented in Section 2.1.1.

Hence, for the CID metric which was used to obtain the subsequent results, image-appearance modeling only consisted of contrast-sensitivity filtering. The following CSFs were chosen for the normalization of the CID metric assuming a visual resolution $VR = 40\,\mathrm{s/deg}$ as suggest in Section 2.2:

1. The chromatic and achromatic CSFs from the *iCAM* framework [117] which were proposed to evaluate image differences. A modification suggested by Johnson and Fairchild [59] was made: the achromatic band-pass CSF is turned into a low-pass filter and clipped above 1. The CSF filtering was done in the frequency domain within the working color space LAB2000HL. The corresponding CID metric is denoted as *CID-CSF1*.

2. The same CSFs were employed in the intensity-linear orthogonal opponent color space *YCC* [117] as well. This normalization refers to *CID-CSF2*.

3. The *S-CIELAB* model [156] also provides CSFs which were applied in the intensity-linear opponent color space AC_1C_2. The convolution of the images within this color space was performed in the spatial domain. The CID metric based on this method is called *CID-CSF3*.

4. To investigate the impact of normalization on the performance of the CID metric, we proposed a fourth version without contrast-sensitivity filtering named *CID-None*.

5.4.3 Training of the CID Parameters

The remaining parameters of CID which have not yet been fixed are the parameters c_1,...,c_5. Each of these parameters corresponds to an image-difference comparison defined in Section 5.3.2 and therefore – according to Equation 5.7 – to one of the IDFs ($\mathbf{L}_L, \mathbf{C}_L, \mathbf{S}_L, \mathbf{L}_C, \mathbf{L}_H$). In particular, c_1 corresponds to the lightness-difference comparison l_L (see Equation 5.8) and thus to $\mathbf{L}_L$, c_2 corresponds to the lightness-contrast comparison c_L (see Equation 5.10) and thus to $\mathbf{C}_L$, c_3 corresponds to the lightness-structure comparison s_L (see Equation 5.11) and thus to $\mathbf{S}_L$, c_4 corresponds to the chroma-difference comparison l_C (see Equation 5.12) and thus to $\mathbf{L}_C$, and c_5 corresponds to the hue-difference comparison l_H (see Equation 5.14) and thus to $\mathbf{L}_H$.

The CID parameters c_1,...,c_5 were determined by training on an IQA database. Most IQA databases involve conventional distortions like noise or blur (see Section 4.3). Thus, an IDM trained on these databases is supposed to have a good prediction performance on such distortions. Conventional distortions occur mainly on the lightness channel which is sufficient for training of IDMs working on grayscale images. However, the CID metric possesses also chromatic channels which need color information for training. Therefore, color-distortion IQA databases – in particular gamut-mapping databases (see Section 4.4) – should be used for training the CID metric.

There are not many publicly available gamut-mapping databases so we chose the *Combined Gamut-Mapping Database* described in Section 4.5 for our investigations. In addition to *training* of the CID parameters, we wanted to *test* the CID metric with respect to the trained parameters. That is why we divided the database into two disjoint sets: the *Combined Gamut-Mapping Training Set* and the *Combined Gamut-Mapping Test Set*. About 50 % of the distorted images from each dataset included in the Combined Gamut-Mapping Database were randomly selected for the training set resulting in 144 reference images, 1 230 corresponding distorted images, and 14 239 observer evaluations. The remaining images were used for test purposes to analyze the CID metric. Properties of the test set are the majority hit rate $\hat{p}_m = 0.795$ and a significant hit-rate difference for differences of 0.01 and higher with respect to a significance level $\gamma = 0.05$.

Training the CID parameters on an IQA database was not a simple task because five variables $(c_1, ..., c_5)$ had to be adjusted. In particular, the hit rate on the Combined Gamut-Mapping Training Set was maximized by varying the CID parameters. Finding a set of CID parameters which results in the maximal hit rate was tedious. For any set of parameters, a computation of the CID metric for each distorted image of the training set and a comparison with all subjective evaluations was required. For the optimization problem, I employed the *patternsearch* MATLAB function which uses a brute force method rather than a gradient-based method. Please note that the result depends on the starting parameter values and it is not guaranteed to find the global hit-rate maximum. The starting parameters are derived

Table 5.1: Parameters of the CID Metric Trained on the Combined Gamut-Mapping Training Set.

	c_1	c_2	c_3	c_4	c_5
CID-CSF 1	0.004	0.538	0.381	0.002	0.006
CID-CSF 2	0.004	0.138	0.350	0.002	0.003
CID-CSF 3	0.002	0.125	0.119	0.003	0.008
CID-None	0.005	0.388	0.425	0.003	0.002
Proposed (valid for all CSFs)	0.002	0.1	0.1	0.002	0.008

by a rough parameter estimation. The results of the CID parameter training for the proposed CSF filterings are summarized in Table 5.1. An example of the multi-scale CID metric with optimized parameters without image normalization (CID-None) is shown in Figure 5.6.

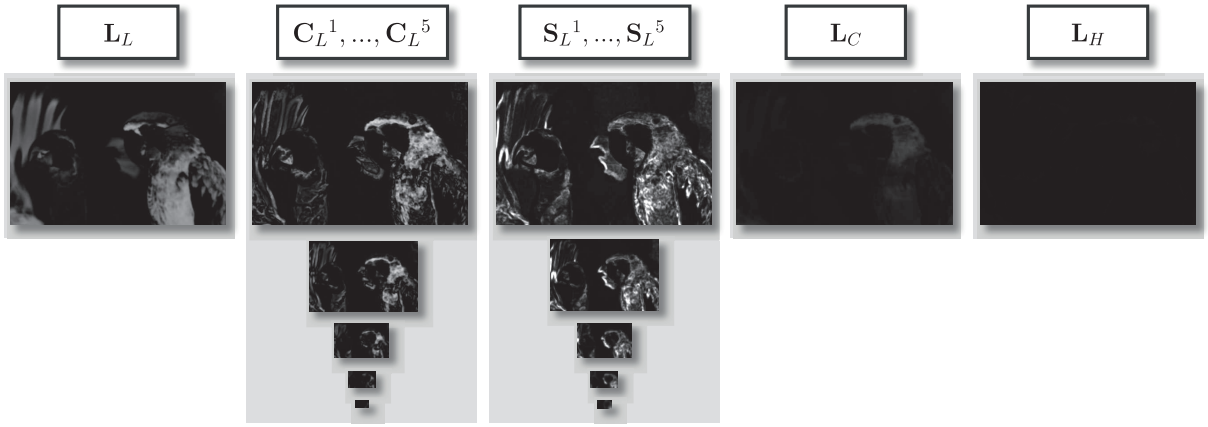


Figure 5.6: All image-difference maps computed by the CID metric without image normalization from the example given in Figure 5.4 (upper row). The corresponding original image as well as the distorted image is taken from the Dugay Database [31]. The original version of the figure has already been published in [76].

The trained parameters are insensitive to small deviations, i.e., deviations of a few thousandths for c_1, c_4, c_5 and deviations of a few tenths for c_2, c_3 do not affect significantly the hit rate on the test set. Hence, we were seeking a unified parameter set valid for each CSF filtering which led to an optimized hit rate. We proposed a parameter set given in Table 5.1 which fulfills the requirement of a simple parameter set without a significant hit-rate difference to the optimized parameter set for each CSF filtering. Nonetheless, in the upcoming sections the results are based on the individually optimized parameters.

The proposed CID parameters show different magnitudes for the parameters c_2 and c_3 of the contrast and structure IDFs **C** and **S** on the one hand and for the parameters c_1, c_4, and c_5 of the difference IDFs **L** on the other hand due to their different constitution. $c_2 = c_3$ indicates an equal weighting of lightness-contrast and lightness-structure differences. Compared with both c_1 and c_4, the parameter c_5 is considerably greater which puts more weight on hue differences rather than on lightness or chroma

differences. This coincides with the constraint of most gamut-mapping algorithms that hue has to be preserved [88].

In [113] I used a different approach of parameter training called *cross validation*. Within the training set I randomly chose 20 % of all reference images and optimized the parameters by maximizing the hit rate on the corresponding subjective evaluations. Then, the hit rate on the remaining evaluations is computed with the resulting parameters. This was done several times, in particular 81 times. By taking the median parameters of their distribution I determined the parameters of the metric. This approach has the advantage that overfitting the data can be detected, a distribution of parameters is attained which allows further analysis, and the optimization on the reduced training set is faster. However, these advantages were rather small for the context of training the CID parameters and the determination of the parameters is not unambiguous.

5.5 Results and Discussion

In the upcoming section, the performance of the CID metric with its optimized parameters is analyzed on the Combined Gamut-Mapping Test Set as well as on the TID2008 database (see Section 4.3.2) including conventional distortions. The different image normalizations as well as the impact of the multi-scale approach are investigated. Due to its good prediction performance on conventional distortions and as the basis for the proposed CID metric, the SSIM index as well as the MSSIM index serve as baseline IDMs.

5.5.1 Impact of Image-Difference Features on Color Distortions

The CID metric proposed in Equation 5.16 is composed of several individual IDFs (see Section 5.3). To analyze the impact of each IDF on the prediction performance, hit rates are computed for all possible combinations of IDFs on a color-distortion IQA database. Only IDFs computed on the first scale (thus, on the entire image) are used. The results are shown in Figure 5.7 for all image normalizations of CID on the test set of the Combined Gamut-Mapping Database. Please note that the parameters referring to an individual IDF for each image normalization were optimized for the combination of all five IDFs (given in Table 5.1).

The performance of an IDF combination is the better, the higher the hit rate $\hat{p}$ which is the fraction of right predictions (see Section 4.2.2). Here, two hit rates are significantly different if their difference is equal to or larger than about 0.01 assuming a significance level $\gamma = 0.05$ (see Section 4.2.4). The combination of IDFs is factorial as for the CID metric (see Equation 5.16), e.g., the combination of the lightness-difference and the lightness-contrast IDF (6th column) is $1 - \mathbf{L}_L \cdot \mathbf{C}_L$. As a comparison, the hit rate of SSIM (see Section 3.3.2) is marked by a red line ($\hat{p} = 0.650$). For a fair comparison, the parameters

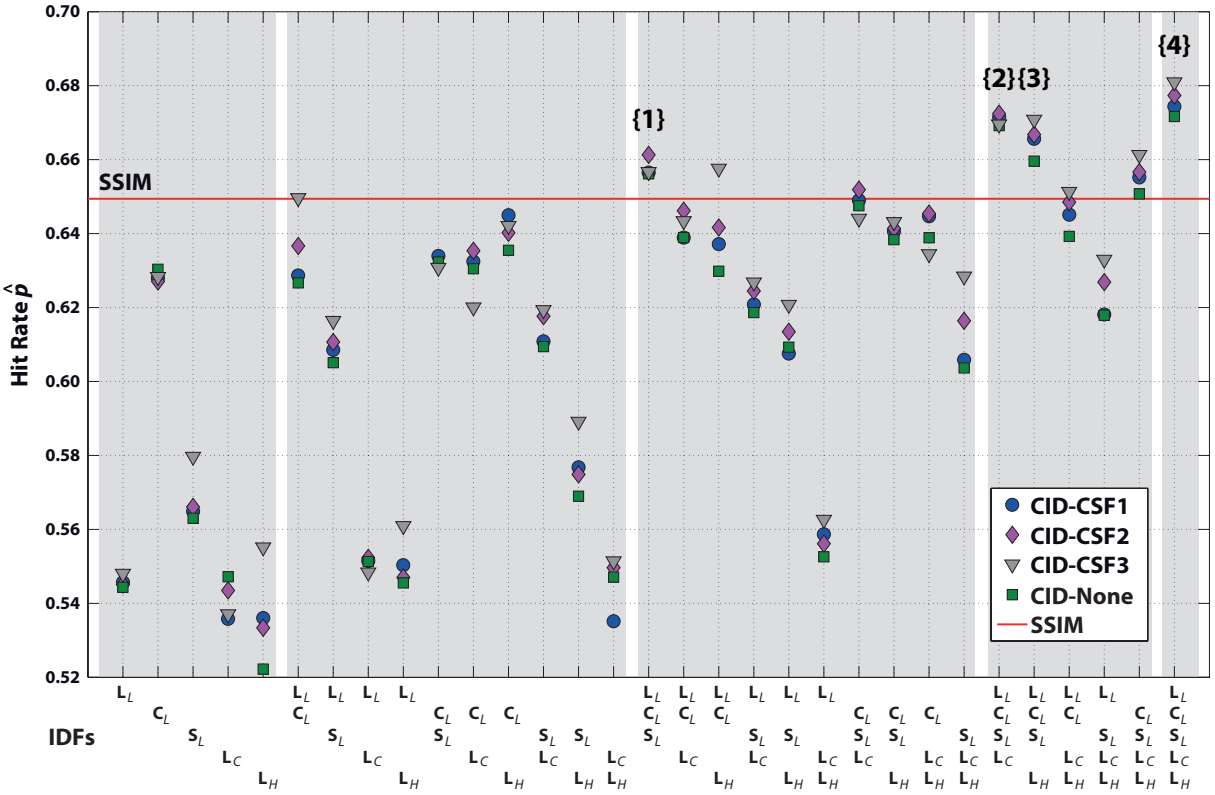


Figure 5.7: Hit rates $\hat{p}$ for all possible combinations of CID's IDFs on the test set of the Combined Gamut-Mapping Database (introduced in Section 4.5). A hit-rate difference equal to or larger than about 0.01 is significant assuming a significance level $\gamma = 0.05$. SSIM's hit rate $\hat{p} = 0.650$ is marked by a red solid line. IDF combinations of interest are labeled by the numbers {1} - {4}. The majority hit rate $\hat{p}_m = 0.795$ is not illustrated. The original version of the figure has already been published in [76].

$c_i, i \in \{1, 2, 3\}$ of the SSIM index were also trained on the Combined Gamut-Mapping Training Set but did not yield a significant difference to the default parameters ($\hat{p} = 0.649$).

Analyzing Figure 5.7 allows some conclusions; the most important are given in the following:

- In most of the cases the difference between the CSF filterings (CID-CSF1, CID-CSF2, CID-CSF3, and CID-None) is not significant such that no particular CSF filtering can be recommended. However, neglecting the visual resolution (CID-None) results in worse hit rates in most of the cases.

- Combining the three achromatic IDFs ($\mathbf{L}_L$, $\mathbf{C}_L$, and $\mathbf{S}_L$; see {1} in the figure) performs better but not significantly better than SSIM – except CID-CSF2 (in particular, SSIM's hit rate $\hat{p} = 0.650$ demands a hit rate $\hat{p} \geq 0.660$ for a significantly better performance). The lightness-difference IDF $\mathbf{L}_L$ adjusted to a nearly perceptually uniform color space (see Equation 5.8) seems to slightly increase the prediction performance compared with the SSIM luminance (see Equation 3.10).

- The most important IDF is lightness-contrast $\mathbf{C}_L$ because adding it to any IDF combination always significantly improves the hit rate.
- The majority hit rate $\hat{p}_m = 0.795$ is still far above the best achieved hit rate $\hat{p} = 0.681$ which is predicted by CID-CSF3 for the combination of all IDFs (see {4} in the figure). I.e., there is much room for improvement which is unlikely to be reached with low-level features disregarding, for instance, image semantics.
- The addition of the chromatic IDFs $\mathbf{L}_C$ (see {2} in the figure), $\mathbf{L}_H$ (see {3} in the figure), or both (see {4} in the figure) to the combination of the achromatic IDFs ($\mathbf{L}_L$, $\mathbf{C}_L$, and $\mathbf{S}_L$) significantly improves the hit rate with respect to the SSIM index. Particularly, CID-CSF3 being the best IDM has an improvement of $\approx 20\%$ on the hit-rate ratio $\hat{p}_r$ (see Section 4.2.3) compared with SSIM: $\hat{p}_r(\text{CID}-\text{CSF3}) = 0.61$ and $\hat{p}_r(\text{SSIM}) = 0.51$.

The combination of all IDFs (see {4} in the figure) refers to the proposed CID metric computed only on the first scale – thus, denoted as *single-scale CID* metric. The performance of CID computed on more than one scale – thus, denoted as *multi-scale CID* metric – is investigated in the next section.

5.5.2 Impact of the Multi-Scale Approach

A multi-scale approach is supposed to lead to a far better prediction performance on conventional distortions, e.g., the multi-scale MSSIM index which uses $m = 5$ scales (see Section 3.3.3) outperforms the single-scale SSIM index on the TID2008 [109]. But does this also apply to gamut-mapping distortions?

To investigate the impact of more than one scale, the hit rate of the CID metric was computed for $m = 1, 2, ..., 5$ scales on the Combined Gamut-Mapping Test Set. The scale weights β_i, $i \in \{1, ..., 5\}$ of the MSSIM index given in Table 3.2 were also used for the CID metric and in case $m < 5$ furthermore normalized to 1. The results for the CID metric as well as for the SSIM (red solid line) and the MSSIM (black dashed line) index are illustrated in Figure 5.8.

Contrary to conventional distortions, the multi-scale approach lowers the prediction performance for gamut-mapping distortions. In particular, the hit rate of MSSIM ($\hat{p} = 0.632$) is significantly lower than for SSIM ($\hat{p} = 0.650$), i.e., applying the multi-scale approach to SSIM – which results in MSSIM – leads to a significant decrease in the hit rate. Nearly the same holds for the CID metric: the multi-scale CID for $m = 4$ or $m = 5$ scales performs worse for all image normalizations (CID-CSF1, CID-CSF2, CID-CSF3, and CID-None) compared to the single-scale CID which corresponds to $m = 1$. For CID-CSF3, the multi-scale approach already significantly lowers the prediction performance of the single-scale CID for $m = 3$ scales.

The discrepancy of the viewing conditions – a visual resolution $VR = 32\,\text{s/deg}$ at the experiment to determine the scale weights $\beta_1, ..., \beta_5$ [151] compared with $VR = 40\,\text{s/deg}$ at the gamut-mapping database

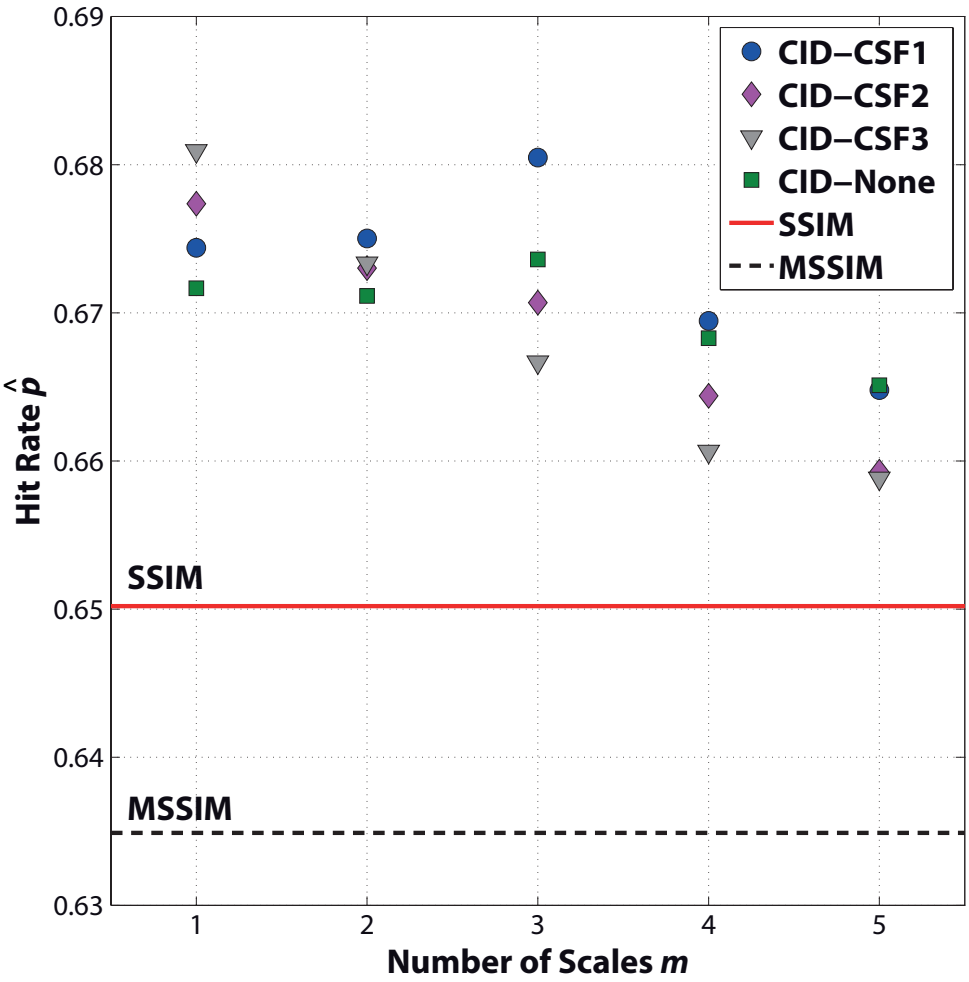


Figure 5.8: Hit-rate comparison of the CID metric dependent on the number of scales m on the Combined Gamut-Mapping Test Set ($m = 1$ corresponds to the single-scale CID). A hit-rate difference equal to or larger than about 0.01 is significant with respect to a significance level $\gamma = 0.05$. SSIM's hit rate $\hat{p} = 0.650$ is marked by a red solid line and MSSIM's hit rate $\hat{p} = 0.632$ by a black dashed line. The majority hit rate $\hat{p}_m = 0.795$ is not illustrated. The original version of the figure has already been published in [76].

– might explain the drop of the multi-scale results. To investigate the discrepancy, we adjusted the scale weights to the database's viewing conditions by interpolating the original parameters. However, the hit rate of the adjusted MSSIM index remains the same (0.632). We therefore concluded that this minor change in the visual resolution is unlikely to have a great influence on the hit rates.

I investigated the negative effect of the multi-scale approach on the gamut-mapping database by checking if lightness distortions from gamut mapping are fundamentally different from conventional distortions. For the factorial combination of lightness-contrast IDF $\mathbf{C}_L$ and lightness-structure IDF $\mathbf{S}_L$, I computed the *Pearson correlation coefficient*[1] between the first scale ($\mathbf{C}_L{}^1\mathbf{S}_L{}^1$) and each single scale,

[1]The *Pearson correlation coefficient* measures a linear correlation between two variables and is also called *Pearson*

$\mathbf{C}_L{}^i\mathbf{S}_L{}^i$, $i = 1, ..., 5$. In Figure 5.9, the correlation is shown for CID and MSSIM on the Combined Gamut-Mapping Test Set as well as on the TID2008 database.

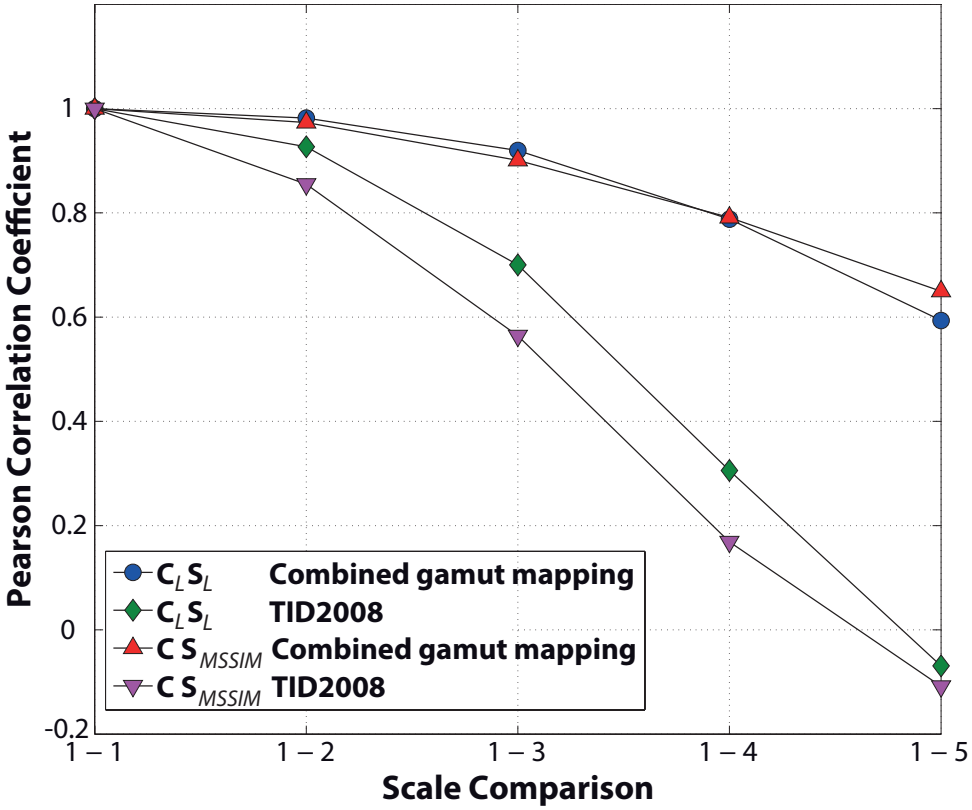


Figure 5.9: Pearson correlation coefficients of the CID lightness-contrast IDF times lightness-structure IDF between the largest scale ($\mathbf{C}_L{}^1\mathbf{S}_L{}^1$) and every single scale ($\mathbf{C}_L{}^i\mathbf{S}_L{}^i$, $i \in \{1, ..., 5\}$) for both the Combined Gamut-Mapping Test Set and the TID2008 database. Analogously the correlations for the MSSIM index. The original version of the figure has already been published in [76].

On gamut-mapping distortions, the scales between the combined IDFs highly correlate between themselves. In other words, the image-difference features extracted from each scale do not vary across scales, i.e., they are rather the same for every scale. The opposite applies to conventional distortions where the scales have a much lower correlation, i.e., the IDFs are rather different for every scale. The same observations hold for the MSSIM index. Indeed, the behavior of gamut-mapping distortions in the lightness component is very different compared with conventional distortions. Finding an appropriate multi-scale approach on gamut-mapping distortions is left to future work.

product-moment correlation coefficient – or simply *correlation coefficient* since it is the most common measure of correlation. It is defined as the covariance divided by the standard deviations of both variables to be compared. The range of the correlation coefficient lies between -1 (totally linear negative correlation) and +1 (totally linear positive correlation). A correlation coefficient of 0 indicates no correlation at all.

Table 5.2: Spearman Correlations on the TID2008 Conventional Distortion Database [76].

	Single-scale	Multi-scale (5 scales)
CID-CSF1	0.643	0.789
CID-CSF1*	0.621	0.797
CID-CSF2	0.642	0.793
CID-CSF2*	0.622	0.790
CID-CSF3	0.750	0.790
CID-CSF3*	0.736	0.798
CID-None	0.579	0.710
CID-None*	0.585	0.792
SSIM	0.625	
SSIM$^+$	0.775	
MSSIM		0.853

*: chromatic image-difference features omitted
$^+$: automatic downsampling

5.5.3 Impact of Color on Conventional Distortions

The performance of the CID metric was also computed on the TID2008 database which represents conventional distortions. Only *mean opinion scores* (MOS, see Section 4.2.1) are provided by TID2008 such that Spearman rank order correlation [135] rather than hit rates were used. The impact of color is analyzed by calculating the CID metric with and without the chromatic chroma-difference and hue-difference IDFs, $\mathbf{L}_C$ and $\mathbf{L}_H$. The multi-scale CID with $m = 5$ scales is computed as well. The results are summarized in Table 5.2 including the SSIM index, SSIM with a reasonable downscaling which corresponds to normalization denoted as SSIM$^+$, and the multi-scale MSSIM which serve as comparison. Please note that the hit-rate difference between SSIM and SSIM$^+$ is not significant on the Combined Gamut-Mapping Test Set but considerably on TID2008.

In the following, some conclusions from Table 5.2 are given:

- If CSF filtering is utilized the hue- and chroma-based IDFs do not considerably affect the accuracy of the CID metric. These IDFs even have a negative influence on the prediction performance if the visual resolution is not considered.

- For single-scale IDMs, the normalization affects the Spearman correlations. The results of the single-scale CID-CSF1 and CID-CSF2 are similar to those of the SSIM index but not as good as those for the S-CIELAB-based CID-CSF3 which almost matches the SSIM$^+$ index.

- Contrary to the gamut-mapping data, the multi-scale approach is beneficial for both the CID metric and MSSIM – compared to the single-scale SSIM and SSIM$^+$. By far the best performance on the

database is the MSSIM index although the multi-scale CID is based on the same approach.

To summarize, the impact of color on conventional distortions is low and the chromatic IDF do not adversely affect the prediction accuracy (except the multi-scale CID-None). The multi-scale CID metric is not as good as the MSSIM index. However, the parameters were trained on gamut-mapping data where the multi-scale approach fails and no conventional distortions are applied. As a trade-off of all results, we proposed the single-scale CID-CSF3 as the default CID metric because for gamut-mapping distortions single-scale leads to significantly better results and single-scale CID's performance is best for S-CIELAB-based filtering on conventional distortions.

5.5.4 Image-Difference Metrics on the Combined Gamut-Mapping Test Set

As already mentioned, there is a lack of investigations of state-of-the-art IDMs on gamut-mapping databases. Hence, I include the hit rates of some renowned IDMs on the Combined Gamut-Mapping Test Set. The results are shown in Table 5.3 – hit-rate differences of 0.01 and larger are significant assuming a significance level $\gamma = 0.05$. The comparison includes the default CID, SSIM (see Section 3.3.2), MSSIM (see Section 3.3.3), FSIM and FSIMc (see Section 3.3.4), PSNR-HVS [33] and PSNR-HVS-M [110], PSNR-HA and PSNR-HMA [106], and the MeTriX MuX package (see Section 3.3.6).

I categorize these metrics into three groups (see Section 3.2.1) which is also shown in Table 5.3: metrics based on structural similarity (*Structure*), metrics based on statistics (*Statistics*), and metrics based on noise fidelity (*Noise*) which is in particular the signal-to-noise ratio (SNR). As can be seen in the table, the metrics which employ the evaluation of structural deviations have the best performance on gamut-mapping databases. The same holds for such metrics on databases with conventional distortions [107] except PSNR-HA and PSNR-HMA which show great performance on TID2013.

The CID metric performs significantly better than all other metrics but is still far below the majority hit rate. Employing color information (right column of the table) is beneficial for CID compared to SSIM as well as for FSIMc compared to FSIM. Color seems to have positive influence on the hit rate on gamut-mapping data as it is supposed to be.

Table 5.3: Hit Rates of Image-Difference Metrics (IDMs) on the Combined Gamut-Mapping Test Set.

IDM	Hit Rate*	Category	Color
Majority Hit Rate	0.795		
CID	0.680	Structure	Yes
SSIM	0.649	Structure	No
FSIMc	0.637	Structure	Yes
MSSIM	0.632	Structure	No
FSIM	0.631	Structure	No
UQI	0.629	Structure	No
VIF	0.592	Statistics	No
PSNR-HMA	0.584	Noise	Yes
PSNR-HA	0.584	Noise	Yes
VSNR	0.581	Noise	No
NQM	0.580	Noise	No
VIFP	0.579	Statistics	No
PSNR-HVS	0.572	Noise	No
PSNR-HVS-M	0.567	Noise	No
MSE	0.565	Noise	No
PSNR	0.565	Noise	No
SNR	0.565	Noise	No
WSNR	0.556	Noise	No
IFC	0.554	Statistics	No

*: hit-rate difference of 0.010 and larger is significant assuming a significance level $\gamma = 0.05$

5.6 Conclusions

A new image-difference framework for assessing full-reference image quality is presented. In a first step, the images are normalized to standard viewing conditions employing an image-appearance model. Then, so-called image-difference features (IDFs) are extracted which are based on hypothesis of perceptually important distortions, especially color distortions from gamut mapping. These features – numerical representations of assumptions on the perception of important achromatic and chromatic distortions – are adopted from the SSIM index. The overall image-difference metric called CID metric is obtained by combining the IDFs.

The framework was tested on the Combined Gamut-Mapping Database that provides the visual resolution as the only viewing condition parameter. Investigations were made on the impact of chromatic IDFs, contrast-sensitivity filtering, and the multi-scale approach adapted from the MSSIM index on the prediction performance on gamut-mapping distortions.

Combining the achromatic IDFs yields a prediction performance not significantly different from the SSIM index. The main conclusion is that an additional chroma-difference IDF and hue-difference IDF significantly improves the accuracy on the database. Therefore, image-difference metrics benefit from including color information. The best performing image-difference metric is the proposed default CID metric which achieves an increase of 20 % on the hit-rate ratio compared with the SSIM index on the Combined Gamut-Mapping Database. The use of chromatic IDFs on conventional distortions from the TID2008 – such as noise and blur – does not adversely affect the prediction accuracy with respect to the TID2008.

Even though the proposed CID metric performs best on the gamut-mapping database, it does not reach the maximal achievable hit rate by far. In other words, there is still room for improving the predictions on gamut-mapping distortions.

The normalization of input images to standard viewing conditions – in this case to the visual resolution – is very important. The investigation of normalization shows that on gamut-mapping distortions as well as on conventional distortions the prediction performance increases with normalization.

The multi-scale approach adopted from the MSSIM index to the lightness-contrast and lightness-structure IDF impairs the prediction accuracy on gamut-mapped images contrary to conventionally distorted images. An analysis of distortions in the lightness channel reveals that IDFs extracted from different scales have a very high inter-scale correlation for gamut-mapping distortions but a very low inter-scale correlation for conventional distortions. Hence, a more suitable multi-scale approach for gamut-mapping distortions may further improve the performance of the CID metric.

Not yet investigated are the images' semantics which are hard to evaluate. E.g., the same distortion is more disturbing in a face than in the background. The *human visual system* (HVS) is still far from being

understood although research in that field has been conducted for decades. An approach of handling semantics is to build formulas which describe one aspect of the HVS. Including models for semantics such as *face detection* is supposed to improve the prediction accuracy of the framework.

The distorted images of the Combined Gamut-Mapping Database are mostly created by gamut-mapping algorithms employing the same heuristics such as reduced chroma and nearly constant hue. Thus, chromatic differences among the distorted images do hardly occur which leads to a misinterpretation of the chromatic channels in the parameter training. A database containing highly uncorrelated color distortions is supposed to provide improved training and test results. Furthermore, to investigate the impact of semantics on the low-level IDFs, both semantic and non-semantic distortions should be inside such a database – as similarly done by Scheller Lichtenauer et al. [122].

5.7 My Contributions

In the following, my contributions to the papers I co-authored and which are cited in this chapter (see Section 5.1) are summarized:

- Investigation of different combining models, in particular the linear, factorial, polynomial, and hybrid model.
- Parameter training on the gamut-mapping database.
- Computation and implementation of the CID metric and its maps.
- Multi-scale analysis referring to Figure 5.9.
- Investigation of the impact of each IDF by creating all possible IDF combinations.

Furthermore, these are my investigations which I presented in this chapter and which have not been published yet:

- Recalculation of the majority hit rate.
- An overview of selected IDMs found in literature on a gamut-mapping database.
- Computation of hit-rate ratios.

6

Optimizing Gamut Mapping With Image-Difference Metrics

The *Color-Image-Difference* (CID) metric introduced in the last chapter shows the best prediction performance on gamut-mapping distortions. CID is therefore supposed to be able to optimize gamut mapping if used as an objective function for minimizing the perceived image difference of the gamut-mapped image to the original image. The structure of this chapter is the following: after introducing gamut mapping as a constrained optimization problem, I propose the CID-based gamut-mapping optimization which shows quite promising results. Since the optimization produces unwanted artifacts, the gamut-mapping optimization and thus the CID metric are improved resulting in the *improved Color-Image-Difference* (iCID) metric. The prediction accuracy of iCID as a metric is verified on gamut-mapping distortions as well as on conventional distortions. In a visual experiment, the iCID-based gamut-mapping optimization significantly outperforms a state-of-the-art spatial gamut-mapping algorithm. The variable iCID metric even allows to define so-called optimization intents referring to the desired color properties of the optimized gamut-mapped image.

6.1 Related Publications

Most of the work introduced in this chapter refers to the journal paper by Preiss et al. [112]:

> J. Preiss, F. Fernandes, and P. Urban. Color-Image Quality Assessment: From Prediction to Optimization. *IEEE Transactions on Image Processing*, 23(3):1366–1378, 2014.

Prior to that, a conference paper as basis for the journal paper was published [115]:

> J. Preiss and P. Urban. Image-Difference Measure Optimized Gamut Mapping. In *IS&T/SID, 20th Color and Imaging Conference*, pp. 230–235, 2012.

This chapter is written in the same way as the previous one: most investigations refer to already published content in the given papers. New content as well as my main contributions are particularly emphasized. To make a clear distinction between my contributions and those of the other co-authors, I use the term *I* for my work and the term *we* whenever other authors were involved.

6.2 Gamut Mapping as a Constrained Optimization Problem

Gamut mapping (see Section 2.4) is still an open field of research due to the complex behavior of the *human visual system* (HVS). Additionally, although showing good prediction performance on selected databases, the CID metric lacks in universality as it is trained on a single gamut-mapping database (see Section 5.4.3). A method to mutually solve both problems is explained in this section.

6.2.1 Gamut Mapping

The common objective of gamut mapping (described in Section 2.4) is to obtain a reproduction which has a minimal perceived image difference to the original image. Therefore, almost all gamut-mapping transformations are based on assumptions on how the HVS evaluates image distortions. The creation of particularly disturbing distortions should be avoided in gamut-mapping algorithms. For instance, the HVS is sensitive to hue shifts when it affects memory colors such as brand or skin colors [139]. Hence, hue is preserved in the majority of the gamut-mapping transformations [88].

Since the beginning of gamut-mapping research, pixel-wise transformations have been investigated. However, not taking into account the images' local contrasts may lead to loss of structure [6]. To preserve local color contrasts, a different approach called *spatial gamut mapping* which is based on local gamut-mapping transformations was recently proposed [6, 42, 84, 89, 158]. An independent comparison of some spatial gamut-mapping transformations was drawn by Bonnier et al. [13].

6.2.2 Constrained Gamut-Mapping Optimization

The property of a perfect *image-difference metric* (IDM) is that its prediction **IDM** totally agrees with the population-average image-difference perception **h**, i.e., $\forall\ \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{I}_{M,N}$:

$$\mathbf{h}(\mathbf{X}, \mathbf{Y}) \leq \mathbf{h}(\mathbf{X}, \mathbf{Z}) \quad \Longleftrightarrow \quad \mathbf{IDM}(\mathbf{X}, \mathbf{Y}) \leq \mathbf{IDM}(\mathbf{X}, \mathbf{Z}) \tag{6.1}$$

where $\mathfrak{I}_{M,N}$ is the set of all sRGB images with M rows and N columns. This means that in a paired-comparison experiment, for instance, the metric mimics the decisions made by the majority of human observers. Please note that in the following, the viewing conditions $\mathcal{P}$ are omitted for the sake of clarity.

So why not using a well-predicting IDM to find the perceived closest reproduction of the original image? This question formulates gamut mapping as a constrained optimization problem employing an IDM which fulfills the property given in Equation 6.1 to a great extent as an objective function:

$$\mathcal{Z} = \underset{\mathbf{Y}\widetilde{\subset}\mathcal{G}}{\operatorname{argmin}} \ \mathbf{IDM}(\mathcal{X}, \mathbf{Y}) \tag{6.2}$$

where $\mathcal{X} \in \mathcal{I}_{M,N}$ is the original image, $\mathcal{Z} \in \mathcal{I}_{M,N}$ is the optimized gamut-mapped image, and the expression $\mathbf{Y}\widetilde{\subset}\mathcal{G}$ means that all colors of image $\mathbf{Y} \in \mathcal{I}_{M,N}$ are within the color gamut $\mathcal{G}$.

So far, solving gamut mapping by minimizing the perceptual difference to the original image has been rarely addressed in literature. Minimizing a metric that is very similar to *S-CIELAB* [156] was proposed by Nakauchi et al. [93]. However, S-CIELAB does only poorly correlate with human judgments in gamut-mapping experiments [13]. A related metric was employed by Kimmel et al. [65] who added gradients to the objective function to preserve local contrast variations. Shortcomings of this method are halo effects – especially along strong edges – and the requirement of a convex color gamut which real devices generally do not possess. Both approaches treat separately the color-channel differences ΔL^*, Δa^*, and Δb^* without considering the direction of difference which may lead, for instance, to adverse hue shifts. Alsam and Farup [3, 4] proposed to use anisotropic diffusion in a related approach. Their method shows promising results as it strongly reduces the halo creation but is not widely used. Fusion of images already gamut-mapped by multiple algorithms was used by Zolliker et al. [157]. The original image is divided into regions and a hue-enhanced modification of the SSIM index chooses the highest rated gamut mapping for each region. Visual experiments revealed that fused images were perceived more similar to the original image than every single gamut-mapped image used for the fusion process. However, since the fused images are built of images from existing gamut-mapping algorithms, the fusion approach is limited by their (local) quality.

6.2.3 Capability of CID As Objective Function

When it comes to gamut-mapped images, the CID metric outperforms all known metrics in approximating the property given in Equation 6.1 (see Section 5.5.4). Therefore, we proposed to use the CID metric as the objective function in Equation 6.2 for optimizing gamut mapping. The methodology of this approach which we called *CID-based gamut-mapping optimization* is explained in detail in the upcoming section.

It is worth mentioning that the parameters of the CID metric are trained on a subset of the Combined Gamut-Mapping Database presented in Section 4.5. This is problematic in that the CID metric is thus inherently biased by heuristics. These heuristics are present in the gamut-mapping database because the gamut-mapping algorithms which generated the database employ these heuristics. For instance, the vast majority of gamut-mapping algorithms preserve hue so that information about the impact of hue

deviations on perceived image differences is missing. Moreover, CID cannot properly assess how sensitive the HVS is to artifacts that have not been created by the algorithms which were used to generate the Combined Gamut-Mapping Database.

The constrained gamut-mapping optimization problem (given in Equation 6.2) is very error-prone if the objective function does not match the property given in Equation 6.1. Shortcomings of the CID metric may create disturbing artifacts during the CID-based gamut-mapping optimization. The other way round, artifacts which appear on such CID-based optimized gamut-mapped images reveal that CID may not be able to assess properly such artifacts. Thus, CID-based optimization might be a method to learn more about CID itself. Moreover, adjusting the CID metric in a way that the optimized images are artifact-free and perceived more similar to the original image is a more general approach than training the metric on a database biased by heuristics.

6.3 Optimizing Gamut Mapping by the Color-Image-Difference (CID) Metric

Due to its modular structure and its well-predicting behavior on gamut-mapping distortions shown in Section 5.5, the CID metric is a promising candidate for the constrained gamut-mapping optimization problem in Equation 6.2. How this is done and how well the CID-based optimization performs is explained in this section.

6.3.1 Modifications of the CID Metric

The CID metric defined in Equation 5.16 needed some modifications to be reasonably used in the proposed CID-based gamut-mapping optimization [112]:

- The proposed gamut-mapping optimization solves Equation 6.2 in a pixel-by-pixel way. To ensure a good reproduction which is valid for all viewing conditions, an image-appearance model must not be applied to predict the perceived image difference. I.e., I refer to CID-None when using the metric for gamut-mapping optimization – this normalization includes only the transformation to the LAB2000HL color space. Otherwise, the results of the optimization may only be optimal for the viewing conditions employed by the normalization.

- Due to the pixel-by-pixel optimization, we omitted the downscaling $\mathbf{D}$ and thus the multi-scale approach.

- Each image-difference comparison should result in values between 0 and 1. Since this is not the case for the lightness-structure comparison s_L (see Section 5.3.2), I took the absolute value of the

covariance $\sigma^L_{\mathbf{xy}}$ to fulfill this requirement. The lightness-structure comparison then reads:

$$s_L(\mathbf{x},\mathbf{y}) = \frac{|\sigma^L_{\mathbf{xy}}| + c_3}{\sigma^L_{\mathbf{x}}\sigma^L_{\mathbf{y}} + c_3}. \tag{6.3}$$

- To exploit locally the modular structure of CID (see Equation 5.16) – i.e., the combination of image-difference features – we chose to first multiply CID's image comparisons pixel-wise and then average over the entire image instead of computing the image-difference features separately and then combining them. Investigations showed that such a modification does not significantly change the prediction performance [76].

- For a local assessment of image differences within the compared images $\mathbf{X},\mathbf{Y} \in \mathcal{I}_{M,N}$, the modified CID metric is defined on any subimage within $\mathbf{X},\mathbf{Y}$. I define all pairs of corresponding windows ($\mathbf{x}$,$\mathbf{y}$) within the subimage as $\mathcal{S} \subseteq \mathcal{T}$ where $\mathcal{T}$ is the set of all pairs of corresponding windows within the entire images $\mathbf{X},\mathbf{Y}$. The smallest possible subimage is a single pixel. Please note that a pixel array $\mathbf{x}$ close to the subimage's edge also contains pixels $\mathbf{x}_i$, $i \in \{1,2,...,k^2\}$ not included in the subimage (k is the side length of the rectangular window).

- In opposite to CID when used as an IDM, the number $|\mathcal{S}|$ of corresponding windows (details about the number of corresponding windows are given in Section 5.3.1) is equal to the number of pixels in the subimage when used for the optimization – this also includes the images' edge pixels. For instance, if CID is computed on the entire images ($\mathcal{S} = \mathcal{T}$), then $|\mathcal{S}| = M \cdot N$ (valid plus virtual windows) for CID used for optimization instead of $|\mathcal{S}| = [M-k+1]\cdot[N-k+1]$ (only valid windows) for CID used for assessing image differences (see padding in Section 5.3.1).

Summarizing the modifications, the CID metric employed as a target function for optimizing gamut mapping and computed on a subimage $\mathcal{S} \subseteq \mathcal{T}$ is defined as [112]:

$$\mathbf{CID}_{\mathcal{S}}(\mathbf{X},\mathbf{Y}) = 1 - \frac{1}{|\mathcal{S}|}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{S}} \left[l_L(\mathbf{x},\mathbf{y})\cdot c_L(\mathbf{x},\mathbf{y})\cdot s_L(\mathbf{x},\mathbf{y})\cdot l_C(\mathbf{x},\mathbf{y})\cdot l_H(\mathbf{x},\mathbf{y})\right]. \tag{6.4}$$

The image-appearance model in the normalization step is omitted, i.e., the pixels $\mathbf{x}_i$, $i = 1,2,...,k^2$ in a pixel array $\mathbf{x}$ directly represent the LAB2000HL values on that pixel position.

As a result, the CID metric now possesses the following property [112]:

$$\left[\mathbf{CID}_{\mathcal{T}}(\mathbf{X},\mathbf{Y}+\mathbf{D}_j) - \mathbf{CID}_{\mathcal{T}}(\mathbf{X},\mathbf{Y})\right]|\mathcal{T}| \;=\; \left[\mathbf{CID}_{\mathcal{W}_j}(\mathbf{X},\mathbf{Y}+\mathbf{D}_j) - \mathbf{CID}_{\mathcal{W}_j}(\mathbf{X},\mathbf{Y})\right]|\mathcal{W}_j| \tag{6.5}$$

where $j \in \{1,...,M\}\times\{1,...,N\}$ is a pixel position within the image, $\mathcal{W}_j$ is a $k \times k$-window centered at j, and $\mathbf{D}_j \in \mathcal{I}_{M,N}$ is an sRGB image with $\mathbf{D}_j(j) \neq (0,0,0)$ (non-zero values at pixel position j)

and $\mathbf{D}_j(l) = (0,0,0)$, $l \neq j$, $l \in \{1,...,M\} \times \{1,...,N\}$ (zero values otherwise). I.e., $\mathbf{Y} + \mathbf{D}_j$ is a pixel deviation of image $\mathbf{Y}$ only at position j and such a pixel deviation affects $\mathbf{CID}_{\mathcal{T}}$ only on a small window $\mathcal{W}_j$ centered at j. In other words, if an image pixel is altered, the CID metric computed on the complete images can be updated by only recomputing the CID metric on a small $k \times k$-window centered at the altered pixel. This property is not valid if CID is computed in the original order (first computing averaged image-difference features and then combining them).

6.3.2 Optimizing Gamut Mapping

The modified CID metric fulfills the requirement to solve the gamut-mapping-optimization problem given in Equation 6.2 in a computationally easy and relatively fast way. I denote this optimization as *CID-based gamut-mapping optimization* which is proposed in the following.

From the property given in Equation 6.5, the following implication is valid $\forall j \in \{1,...,M\} \times \{1,...,N\}$ [112]:

$$\begin{aligned} \mathbf{CID}_{\mathcal{W}_j}(\mathbf{X}, \mathbf{Y} + \mathbf{D}_j) &\leq \mathbf{CID}_{\mathcal{W}_j}(\mathbf{X}, \mathbf{Y}) \\ \Longleftrightarrow \mathbf{CID}_{\mathcal{T}}(\mathbf{X}, \mathbf{Y} + \mathbf{D}_j) &\leq \mathbf{CID}_{\mathcal{T}}(\mathbf{X}, \mathbf{Y}) \end{aligned} \tag{6.6}$$

since $\mathbf{CID}_{\mathcal{S}} \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{T}$. This property even though less constraining than the property given in Equation 6.5 reveals that a descent direction for the CID-based gamut-mapping optimization is reached by changing one pixel so that $\mathbf{CID}_{\mathcal{W}_j}$ is reduced.

We turned the continuous problem in Equation 6.2 into a discrete optimization by quantizing the working color space LAB2000HL. This was done by subdividing the perceptually uniform color space into equally-sized cubes[1] with edges along the independent L, a, and b directions. Exploiting the whole range of the LAB2000HL color space, lightness channel L was encoded in 7-bit and the red-green channel a as well as the blue-yellow channel b were encoded in 8-bit. Each cube represents one color in the quantized space. Since the cube's edge length δ is below the just-noticeable difference (JND)[2], adjacent colors are perceived to be equal within the quantized LAB2000HL color space. For the proposed optimization, this color space is the domain of all pixel colors.

The CID-based gamut-mapping optimization is an iterative, pixel-by-pixel solution to Equation 6.2. Given a color gamut $\mathcal{G}$ and the original image $\mathcal{X}$ to be reproduced, the optimization requires a starting

[1] Actually, LAB2000HL was subdivided into cuboids rather than cubes due to the selected encoding. However, I refer to cube for an easier description without violating the generality. The cuboid's largest edge length is still below the JND.

[2] The cube's edge length in lightness direction is $\Delta L \approx 0.6$ in the LAB2000HL color space which corresponds to $\Delta L^* \approx 0.8$ in the CIELAB color space, in red-green direction $\Delta a \approx 0.4$ corresponding to $\Delta a^* \approx 0.5$, and in blue-yellow direction $\Delta b \approx 0.4$ corresponding to $\Delta b^* \approx 0.5$. This leads to a color difference ${\Delta E^*_{ab}}^{\delta} \approx 1.1$ (ΔE^*_{ab} is defined in Section 3.2.2) which is below the JND (${\Delta E^*_{ab}}^{\mathrm{JND}} \approx 2.3$ after [126]).

image $\mathcal{Y} = \mathbf{G}(\mathcal{X}) \,\widetilde{\subset}\, \mathcal{G}$ transformed by a standard gamut-mapping transformation $\mathbf{G}$. In the first optimization step, image $\mathbf{Y} = \mathcal{Y} \,\widetilde{\subset}\, \mathcal{G}$ is improved with respect to CID by changing one image pixel resulting in an in-gamut image $\mathbf{Z} \,\widetilde{\subset}\, \mathcal{G}$. Then, for each further optimization step set $\mathbf{Y} = \mathbf{Z}$ and improve the image the same way by always changing a different image pixel. One iteration is finished when each image pixel was involved in an optimization step. The CID-based gamut-mapping optimization is completed if a predefined termination condition is fulfilled after an iteration which results in the optimized gamut-mapped image $\mathcal{Z}$.

For each optimization step, the improved reproduction $\mathbf{Z}$ of image $\mathbf{Y}$ satisfying $\mathbf{CID}_{\mathcal{T}}(\mathcal{X}, \mathbf{Z}) \leq \mathbf{CID}_{\mathcal{T}}(\mathcal{X}, \mathbf{Y})$ is obtained as follows [112]:

$$\begin{aligned} \mathbf{Z}(j) &= \mathbf{Y}(j) + \underset{\mathbf{D}_j | \mathbf{D}_j(j) \in \mathcal{H}(j)}{\operatorname{argmin}} \mathbf{CID}_{\mathcal{W}_j}(\mathcal{X}, \mathbf{Y} + \mathbf{D}_j) \\ \mathbf{Z}(l) &= \mathbf{Y}(l), \quad \forall l \in \{1, ..., M\} \times \{1, ..., N\},\ l \neq j \end{aligned} \tag{6.7}$$

where $j \in \{1, ..., M\} \times \{1, ..., N\}$ is the pixel position of the optimizing step and

$$\mathcal{H}(j) = \left\{ \epsilon \in \{-\delta, 0, \delta\}^3 \mid \mathbf{Y}(j) + \epsilon \in \mathcal{G} \right\}. \tag{6.8}$$

CID's local property given in Equation 6.5 allows to improve image $\mathbf{Y}$ locally around pixel position j. The methodology of the CID-based gamut-mapping optimization is shown in Figure 6.1.

The color of the improved pixel j stays within the color gamut due to the definition of $\mathcal{H}(j)$. Furthermore, the quantized color space ensures that $\mathbf{CID}_{\mathcal{W}_j}$ is computed only up to 27 times (the maximal number of elements in $\mathcal{H}(j)$) per optimization step. Hence, a brute-force computation is considered as $\mathbf{CID}_{\mathcal{W}_j}$ and $\mathcal{H}(j)$ allow a very efficient performance:

- $\mathbf{CID}_{\mathcal{W}_j}$ does not need to be evaluated for colors $\mathbf{Y}(j) + \epsilon \notin \mathcal{G}$ outside the color gamut.
- $\mathbf{CID}_{\mathcal{W}_j}$ has to be computed for each element of $\mathcal{H}(j)$ – except outside-gamut colors – only for the first iteration. Then, these values are known and can be reused.
- Thus beginning with the second iteration, certain image-difference comparisons are already computed if the deviation ϵ is zero within the corresponding color channels. Moreover, element $\{0\}^3 \in \mathcal{H}(j)$ does not need to be considered any more.
- Some image-difference comparisons stay constant for certain elements of $\mathcal{H}(j)$, e.g., the lightness-based comparisons (l_L, c_L, s_L) only need to be evaluated two times for lightness deviations $\pm\epsilon$.
- The non-constant image-difference comparisons are derived by mathematically simple and computationally inexpensive updates. Only one element of each Gaussian-weighted mean (i.e., sum) needs to be interchanged including at worst square roots.

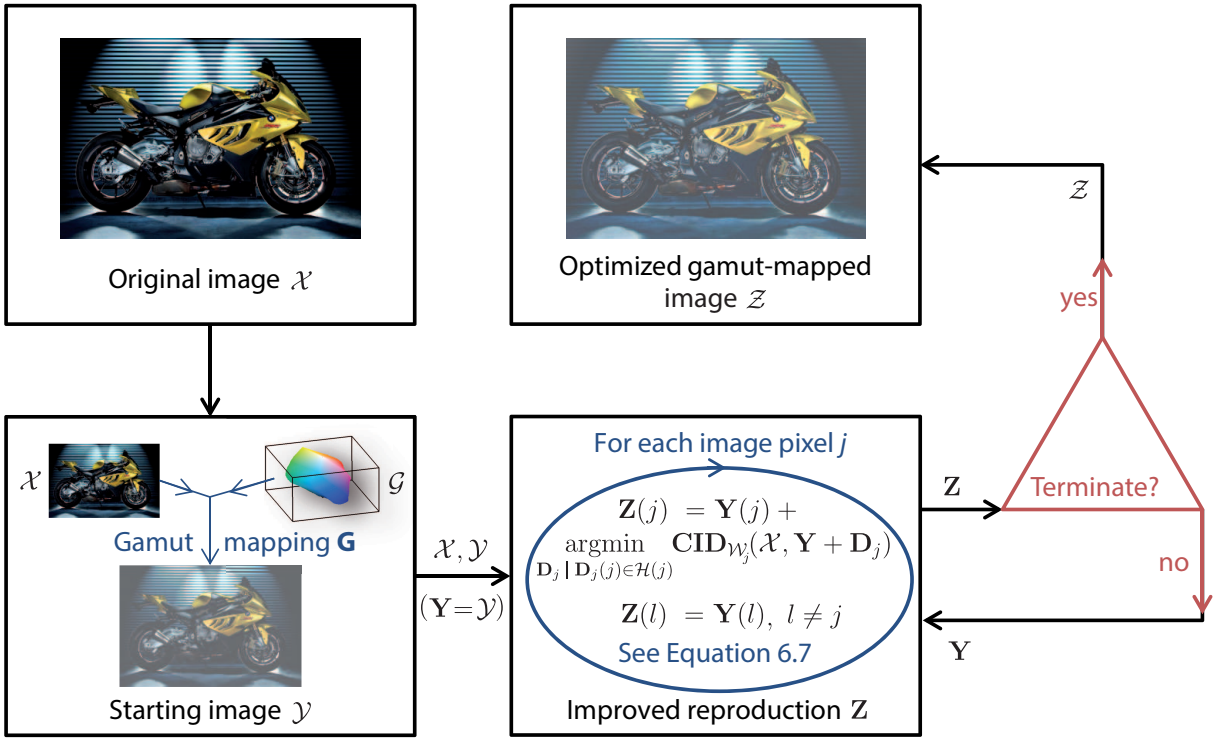


Figure 6.1: Methodology of the proposed CID-based gamut-mapping optimization. Given the original image $\mathcal{X}$ and a color gamut $\mathcal{G}$, first the starting image $\mathcal{Y}$ is obtained by a standard gamut-mapping transformation. Then, CID is minimized for each image pixel j according to Equation 6.7 (one iteration). When the termination condition is fulfilled, the optimization is completed resulting in the optimized gamut-mapped image $\mathcal{Z}$. The original version of the figure has already been published in [112].

- Parallel computing is possible by executing concurrent image-pixel optimizations whose surrounding $k \times k$-windows do not mutually overlap.

The computation of the discrete optimization by applying Equation 6.7 is given by a simplified outline in **Algorithm 1**.

Algorithm 1 CID-Based Gamut-Mapping Optimization

INPUT: `color gamut` $\mathcal{G}$, `original image` $\mathcal{X}$

1. $\mathbf{Y} = \mathbf{G}(\mathcal{X}) \,\widetilde{\subset}\, \mathcal{G}$
2. **REPEAT**
3. **FOR EACH** $j \in \{1, ..., M\} \times \{1, ..., N\}$
4. $\mathbf{Y}(j) = \mathbf{Y}(j) + \operatorname{argmin}_{\mathbf{D}_j \mid \mathbf{D}_j(j)\in\mathcal{H}(j)} \mathbf{CID}_{\mathcal{W}_j}(\mathcal{X}, \mathbf{Y}+\mathbf{D}_j)$
5. **END FOR**
6. **UNTIL** `TERMINATE`

OUTPUT: `optimized gamut-mapped image` $\mathbf{Y}$

Please note that the iterative approach does not guarantee to find the global minimum of Equation 6.2 due to the numerous degrees of freedom. However, the optimization converges to a local minimum because the image-difference prediction of CID is constrained to be decreasing for each pixel deviation. The optimization's step length δ corresponding to the distance of two adjacent pixels in the quantized working color space is a fixed fraction of the JND. Thus, each pixel deviation leads to a result which is perceptually equal to a result with an even smaller step length. Moreover, the result of the iterative process strongly depends on the starting image which determines into which minimum the optimization runs.

6.3.3 Evaluation of CID-Based Gamut-Mapping Optimization

In a first investigation [115], the proposed gamut-mapping optimization was evaluated on five test images (denoted as *parrots*, *honey bee*, *Lula*, *lighthouse*, and *speedway*) using a very small color gamut extracted from the `USNewsprintSNAP2007.icc` profile. To get the starting images $\mathcal{Y}$, two distinct common gamut-mapping algorithms $\mathbf{G}$ were applied on the original images $\mathcal{X}$: *CLIPSLIN* (lightness and chroma clipping in direction of the middle-gray value) and *SGCK* (sigmoidal lightness mapping dependent on chroma followed by knee scaling in direction of the CUSP) [88,115]. The CID-based gamut-mapping optimization was performed on the starting images resulting in ten optimized gamut-mapped images $\mathcal{Z}$.

As a termination condition, we used a fixed number of 30 iterations to finish the optimization. Another condition – which was not considered here – could terminate the optimization if the $\mathbf{CID}_{\mathcal{T}}$ difference between two succeeding iterations falls below a threshold. As expected, the metric $\mathbf{CID}_{\mathcal{T}}$ computed on the entire images decreases after each optimization step and iteration. Already after 10 to 20 iterations, the decrease of $\mathbf{CID}_{\mathcal{T}}$ is relatively small after each iteration. Figure 6.2 shows the behavior of $\mathbf{CID}_{\mathcal{T}}(\mathcal{X}, \mathcal{Z})$ versus variable number of iterations for the five test images and for both gamut-mapping algorithms (CLIPSLIN and SGCK).

In a visual paired-comparison experiment, each optimized image was shown together with the corresponding starting and original image. The original image $\mathcal{X}$ was centered on a calibrated display and starting image $\mathcal{Y}$ as well as optimized image $\mathcal{Z}$ were placed left and right to it in random order. 13 unbiased color-normal observers were asked to choose the best reproduction. Each image comparison was shown twice in reverse left-right order.

As a result, 82.6 % preferred the optimized images. In case of CLIPSLIN, 87.1 % selected the optimized images and 78.2 % in case of SGCK. The optimized images are significantly better than the common gamut-mapped images with respect to a significance level of $\gamma = 0.01$. Please note that we did not compare images based on CLIPSLIN to images based on SGCK.

Even though the results for CID-based gamut-mapping optimization are quite promising, the optimized images contain visible artifacts. In the visual experiment, the gain in contrast and structure of the

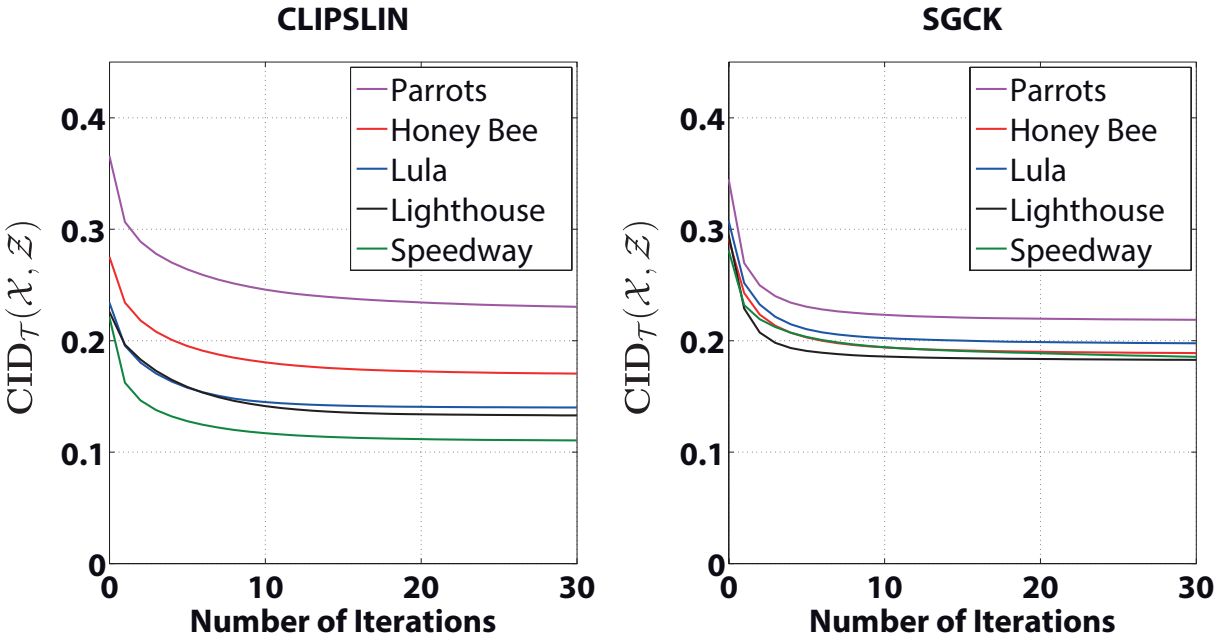


Figure 6.2: Predicted image difference $\mathbf{CID}_{\mathcal{T}}$ between original $\mathcal{X}$ and optimization $\mathcal{Z}$ versus the optimization's iteration number. Computed for five test images and two gamut-mapping algorithms (CLIPSLIN left, SGCK right). The original version of the figure has already been published in [115].

optimized images has a larger impact on the perceived image difference than the apparent artifacts. However, artifacts are not acceptable in gamut mapping. This problem has to be addressed if the proposed gamut-mapping method is desired in real applications.

In particular, the artifacts include lightness inversion, chromatic ringing, chromatic edges, and lightness banding. They are created by the CID-based gamut-mapping algorithm while the predicted image difference to the original image is reduced noticeably. This is in contradiction with Equation 6.1 indicating that CID judges those artifacts to be less important than other distortions. How the shortcomings of the CID metric and the gamut-mapping optimization – in particular artifacts – are handled is explained in the following section.

6.4 Improving the Color-Image-Difference (CID) Metric

Since some artifacts occur at the gamut-mapping optimization, predictions of the CID metric were not always consistent with human perception. By analyzing those artifacts, the CID metric can be improved by adjusting and further modifying it. Artifact-free optimization results are obtained by addressing each artifact and proposing modifications of the metric as listed below.

6.4.1 Lightness Inversion

Lightness inversion describes an inverse relation of two neighboring colors in the lightness channel, e.g., bright-dark turns into dark-bright. An example of the artifact is illustrated in Figure 6.3: the pin-stripes are brighter than the background in the reference image but darker in the optimized image. Lightness inversion is detectable by the CID metric's lightness-difference comparison (see Equation 5.8), i.e., the predicted difference to the reference image is lower for an image without lightness inversion compared to the same image owning this artifact. Hence, it is unlikely that during the gamut-mapping optimization lightness inversion is created. But if the starting image already possesses lightness inversion the effect is preserved or even amplified by the optimization. This is due to the fact that the lightness-contrast comparison and the lightness-structure comparison do not detect lightness inversion which does not affect contrast and structure. In the step-wise below-JND optimization, it is not possible to resolve lightness inversion while keeping lightness-contrast and lightness-structure. So the optimization runs into a local minimum if lightness inversion occurs at the starting image. Analyzing images with lightness inversion confirms that the starting images already possessed lightness inversion from the initial gamut mapping and that it is not created by the optimization.

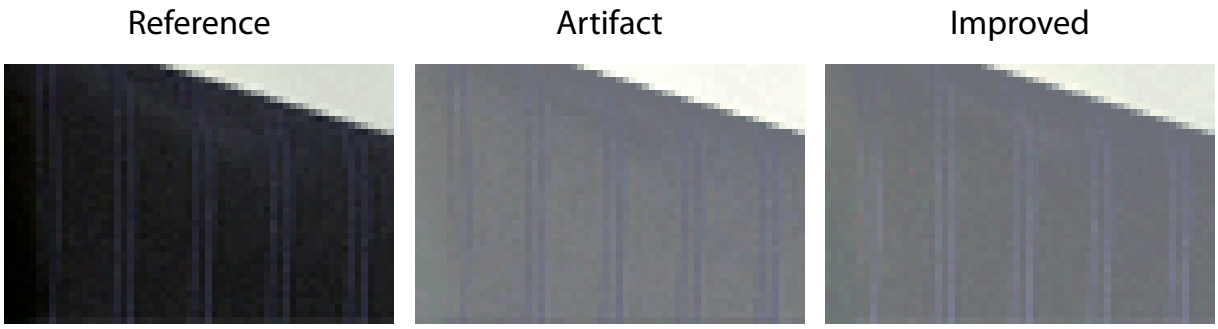


Figure 6.3: Lightness inversion artifact of CID-based gamut-mapping optimization (middle column) and resolved artifact by the improved gamut-mapping optimization (right column). The original version of the figure has already been published in [112].

Therefore, the solution of lightness inversion is an appropriate initial gamut mapping which transforms lightness monotonically, i.e., for original image $\mathcal{X}$ and starting image $\mathcal{Y} = \mathbf{G}(\mathcal{X})$: $L_{\mathcal{X}_i} \leq L_{\mathcal{X}_j} \Rightarrow L_{\mathcal{Y}_i} \leq L_{\mathcal{Y}_j}$ where $L_{\mathcal{X}}, L_{\mathcal{Y}}$ are the lightness components of the images $\mathcal{X}, \mathcal{Y}$ and $i, j \in \{1, ..., M\} \times \{1, ..., N\}$ are arbitrary pixel positions. For the improved optimization, we proposed to apply the gamut-mapping transformations incorporated in ICC profiles as initial gamut mapping because they usually satisfy this condition.

6.4.2 Chromatic Ringing

Chromatic ringing is an artifact caused by chromatic lines which repeat the shape of an image's object several times as can be seen in Figure 6.4. An analysis of the artifact reveals that chromatic ringing exists only in the chroma channel and not in the hue channel. Even though the chroma-difference comparison of the CID metric (see Equation 5.12) accounts for chroma deviations, it is not able to detect chromatic ringing. The comparison evaluates only weighted average chroma values rather than chroma-contrast variations which leaves too many degrees-of-freedom for chroma in the optimization.

Figure 6.4: Chromatic ringing artifact of CID-based gamut-mapping optimization (middle column) and resolved artifact by the improved gamut-mapping optimization (right column). Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The original version of the figure has already been published in [112].

As a solution to chromatic ringing, I introduced the *chroma-contrast* comparison c_C – analogously to the lightness-contrast comparison c_L (see Equation 5.10) but using chroma instead of lightness – as a further image-difference comparison:

$$c_C(\mathbf{x},\mathbf{y}) = \frac{2\sigma_{\mathbf{x}}^C\sigma_{\mathbf{y}}^C + c_6}{{\sigma_{\mathbf{x}}^C}^2 + {\sigma_{\mathbf{y}}^C}^2 + c_6} \tag{6.9}$$

where parameter $c_6 > 0$ and $\sigma_{\mathbf{x}}^C, \sigma_{\mathbf{y}}^C$ are the Gaussian-weighted standard deviations σ_x, σ_y (see Equation 3.8) computed on the chroma channel C. Again, the influence of contrast masking – but on the chroma channel – is set by parameter c_6. After appending the chroma-contrast comparison c_C factorially to the CID metric, no more chromatic ringing occurred during the optimization.

6.4.3 Chromatic Edges

The artifact of contouring the objects' edges by a thin chromatic line is called *chromatic edges*. Figure 6.5 shows an example of chromatic edges at the transitions to the blue areas. Even though chromatic edges occur only in the chroma channel, they do not affect neither the chroma-difference nor the chroma-contrast comparison.

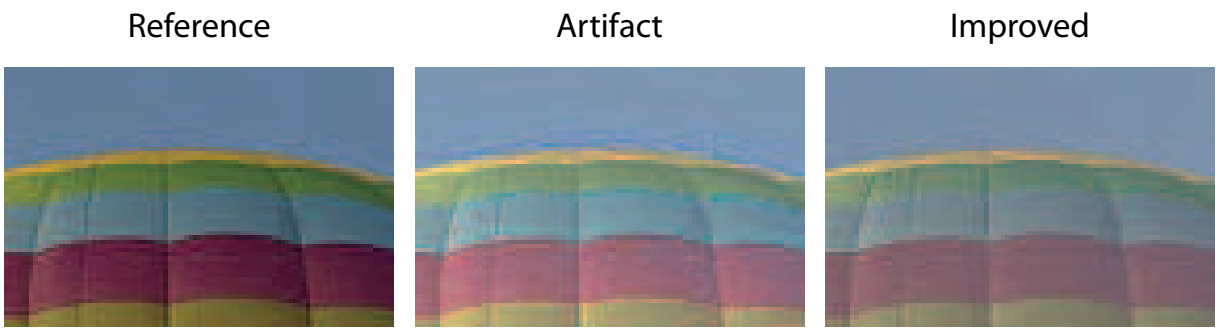


Figure 6.5: Chromatic edges artifact of CID-based gamut-mapping optimization (middle column) and resolved artifact by the improved gamut-mapping optimization (right column). Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The original version of the figure has already been published in [112].

However, chromatic edges bias the structure in the chroma channel of the reference image. Therefore, I proposed the *chroma-structure* comparison s_C which is defined analogously to lightness-structure comparison s_L (see Equation 5.11) on the chroma channel:

$$s_C(\mathbf{x}, \mathbf{y}) = \frac{|\sigma^C_{\mathbf{xy}}| + c_7}{\sigma^C_{\mathbf{x}}\sigma^C_{\mathbf{y}} + c_7} \tag{6.10}$$

where parameter $c_7 > 0$ and $\sigma^C_{\mathbf{xy}}$ is the Gaussian-weighted covariance σ_{xy} (see Equation 3.9) computed on the chroma channel C. The chroma-structure comparison s_C was added to the CID metric yielding optimized gamut-mapped images without chromatic edges.

6.4.4 Lightness Banding

Elongated structures which do not appear in the reference image are called banding artifacts. During the optimization, banding occurs only in the lightness channel and is therefore called *lightness banding*. An example is shown in Figure 6.6. Even though lightness banding is detected by the lightness-structure comparison s_L (see Equation 5.11), it is still created during the optimization. Obviously, the influence of s_L on the overall CID metric is too small and cannot be sufficiently adjusted by the corresponding parameter c_3.

We suggested to raise the importance of the lightness-structure comparison by an exponent $\alpha > 1$ as $s_L(\mathbf{x}, \mathbf{y})^\alpha$ to avoid lightness-banding artifacts. The concept of exponents to weight the contribution of image-difference comparisons was already proposed for the general SSIM index [149]. A sufficiently large exponent α gained from a parameter study as described below inhibited lightness banding at the gamut-mapping optimization as shown in the right image of Figure 6.6.

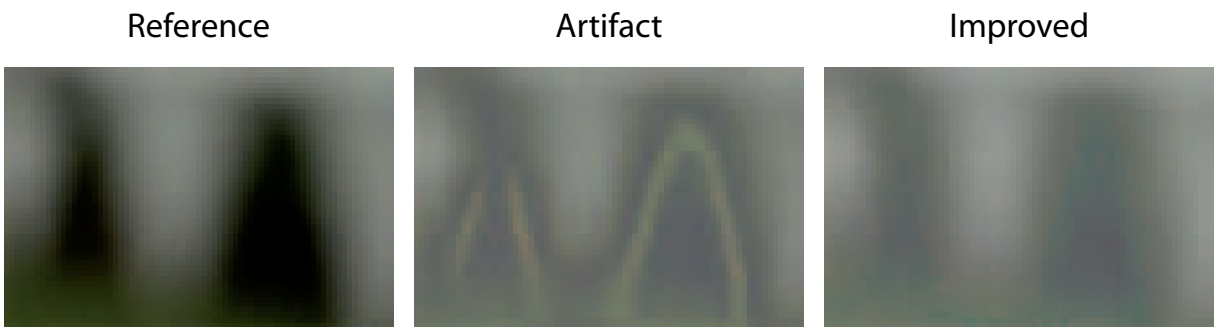


Figure 6.6: Lightness banding artifact of CID-based gamut-mapping optimization (middle column) and resolved artifact by the improved gamut-mapping optimization (right column). The original version of the figure has already been published in [112].

6.4.5 The Improved CID (iCID) Metric

Including all modifications to the CID metric in Equation 6.4 results in the *improved Color-Image-Difference* (iCID) metric:

$$\begin{aligned}&\mathbf{iCID}_{\mathcal{S}}(\mathbf{X},\mathbf{Y}) = \\ &1-\frac{1}{|\mathcal{S}|}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{S}}\left[l_L(\mathbf{x},\mathbf{y})\cdot c_L(\mathbf{x},\mathbf{y})\cdot s_L(\mathbf{x},\mathbf{y})^{\alpha}\cdot l_C(\mathbf{x},\mathbf{y})\cdot l_H(\mathbf{x},\mathbf{y})\cdot c_C(\mathbf{x},\mathbf{y})\cdot s_C(\mathbf{x},\mathbf{y})\right]\end{aligned} \tag{6.11}$$

where l_L,c_L,s_L,l_C, and l_H are the image-difference comparisons of the original CID metric (defined in Section 5.3.2), c_C is the newly defined chroma-contrast comparison (see Equation 6.9), s_C is the newly defined chroma-structure comparison (see Equation 6.10), $\alpha > 1$ increases the importance of the lightness-structure comparison, and $\mathcal{S}$ is defined as in Equation 6.4. As for the CID metric, the input images $\mathbf{X}, \mathbf{Y}$ are first normalized to standard viewing conditions by transformation $\mathbf{N}$ (see Equation 5.3) into the working color space LAB2000HL. All pairs of corresponding windows $(\mathbf{x}, \mathbf{y})$ are extracted from the subimage $\mathcal{S}$ within the normalized images $\hat{\mathbf{X}}, \hat{\mathbf{Y}}$.

6.4.6 Parameter Adjustment

Excluding parameters concerning the viewing conditions, the iCID metric has a total of ten parameters: seven image-difference-comparison parameters $(c_1, ..., c_7)$, lightness-structure weighting exponent α, sliding window side length k, and standard deviation σ_G of the circular-symmetric Gaussian weighting w. CID's parameters were determined by training on a gamut-mapping database which employs certain heuristics as described in Section 6.2.3. However, these highly biased parameters produce artifacts at the gamut-mapping optimization. Here, the parameters of iCID are determined by eliminating the artifacts of the optimization which is a *chicken-and-egg problem*: the gamut-mapping optimization requires a well-predicting IDM to produce artifact-free results and the results of the optimization belonging to a specific

parameter set need to be evaluated by a well-predicting IDM due to the large number of parameters.

However, the CID metric yields the best performance on gamut-mapping distortions (see Section 5.5.4) so that visual judgments are required to determine the parameters. To reduce the number of parameters, some parameter values can be assumed to be equal: the almost perceptually uniform LAB2000HL color space suggests to weight lightness, chroma, and hue equally. Image-difference comparisons with a similar structure can therefore be assumed to have the same parameters. This is the case for lightness-difference comparison, chroma-difference comparison, and hue-difference comparison, i.e., $c_1 = c_4 = c_5$. The image-appearance-model transformation $\mathbf{N}$ (see Equation 5.3) takes into account the different contrast sensitivities for lightness, chroma, and hue of the human visual system. Assuming an appropriate blurring of the input images' color channels at the normalization step and perceptual uniformity of the working color space, lightness-contrast comparison and chroma-contrast comparison can be set equally, i.e., $c_2 = c_6$. Because the same holds for lightness-structure comparison and chroma-structure comparison, we set $c_3 = c_7$. Due to the similarity of contrast comparisons and structure comparisons, they are set equally to further reduce the number of effective parameters, i.e., $c_2 = c_6 = c_3 = c_7$. Please note that exponent α additionally adjusts the relative importance of the lightness-structure comparison. Summarizing, the seven parameters $c_i, i = 1, ..., 7$ are reduced to only two effective parameters, i.e., $c_1 = c_4 = c_5$ and $c_2 = c_6 = c_3 = c_7$.

We kept the sliding window side length $k = 11$ as for the preceding CID metric. This is justified because we assumed viewing conditions with a visual resolution $VR = 40$ samples (i.e., pixels) per degree of visual angle (see Section 2.2): for $k = 11$, the window covers spatial frequencies SF (see Section 2.1.1) in the range of about four to twenty cycles per degree of visual angle. Please note that for one cycle at least two pixels are necessary. This range includes – for a typical display luminance – the peak contrast sensitivity at a spatial frequency of around eight cycles per degree [80]. On the one hand, for smaller k the minimal spatial frequency of the covered frequency range increases. If $k \leq 6$, the peak contrast sensitivity is not covered. On the other hand, for larger k the computational effort is larger because mean μ_x, standard deviation σ_x, and covariance σ_{xy} are weighted sums with $k \times k$ elements (see Section 3.3.2). Therefore, $k = 11$ is a good compromise and in accordance with the original value of the renowned SSIM index [149].

Compared to CID, the standard deviation σ_G of iCID is slightly increased to $\sigma_G = 2.0$. Thus, more weight is given to edge pixels of the window which reduces lightness banding within images resulting from the gamut-mapping optimization. Please note that the combination of k and σ_G is only valid for typical display viewing conditions ($VR = 40$ samples per degree of visual angle). For viewing conditions which do not fulfill this requirement, an appropriate rescaling is necessary. Otherwise, the image scale that iCID is working on does not correspond to the window side length k to which the metric was optimized.

Table 6.1: Adjusted Parameters of the iCID Metric.

c_1, c_4, c_5	c_2, c_3, c_6, c_7	α	k	σ_G
0.002	10	3	11	2.0

A two-stage strategy was used for the adjustment of the remaining effective parameters: $c_1 = c_4 = c_5$, $c_2 = c_6 = c_3 = c_7$, and α. In the first stage, the lightness-structure weighting exponent was set $\alpha = 1$. The remaining parameters' values were first estimated from the CID values and then varied logarithmically as follows: $c_1 = c_4 = c_5 = \{0.0002, 0.002, 0.02, 0.2\}$ and $c_2 = c_6 = c_3 = c_7 = \{0.01, 0.1, 1, 10\}$. Then, iCID-based gamut-mapping optimizations were computed for each parameter combination on 14 reference images showing a variety of scenes. Examples of reference images are given in Figures 6.8 and 6.9. We used the small newspaper gamut of the ICC profile `USNewsprintSNAP2007.icc` as the target color gamut. The profile also includes the initial gamut-mapping transformation (ICC perceptual intent [55]).

A logarithmic parameter variation is justified since iCID-based gamut-mapping optimizations are very insensitive to small deviations of the parameters $c_i, i = 1, ..., 7$. This observation is in agreement with the insensitivity of CID's prediction performance for small parameter deviations (see Section 5.4.3).

All results of the variational study were visually inspected by three color normal observers only with respect to artifacts. Mainly lightness banding occurred because other artifacts almost vanished by introducing iCID rather than CID. At the end of the first stage, the parameter combination was chosen so as to show the best behavior regarding lightness-banding artifacts. The other parameter combinations resulted in more artifacts or in artifacts which were perceived more disturbing.

In the second stage, we increased parameter α using natural numbers ($\alpha = 1, 2, ...$) while keeping the other parameters constant. Only natural numbers were considered because they allow an easier model and faster computation. The iCID-based gamut-mapping optimization was computed for each lightness-structure weighting exponent α using the reference images and the final parameter combination from stage one. Finally, the smallest α was selected which yielded artifact-free optimization results.

A third stage where the selected lightness-structure weighting exponent α was taken for a parameter variation study like done in the first stage was not needed. After the second stage the optimized images were already free from artifacts. The final parameters of the iCID metric which are used in the following are summarized in Table 6.1.

I want to stress the peculiarity of the proposed parameter adjustment: instead of *training* on an IQA database – like common approaches do – the parameters are selected so as to yield *visually* artifact-free iCID-based gamut-mapping optimization results. This approach might not be perfect with respect to iCID's prediction performance on specific databases. Nevertheless, the parameters are not biased by any database which speaks for a high generalization ability of the iCID metric. Furthermore, only three

weighting parameters are needed to produce artifact-free optimizations when iCID is used as a target function. These findings were validated on various other color gamuts and reference images.

6.5 Performance of the Improved Color-Image-Difference (iCID) Metric

The promising results of the iCID-based gamut-mapping optimization need to be confirmed by a thorough analysis which is carried out in the upcoming section. On the one hand, the behavior of iCID as a metric is tested on a gamut-mapping as well as on a conventional image-quality-assessment database. Even though iCID has not been fitted to a particular visual database (see Section 6.2.3), its agreement with subjective scores remains the metric's performance indicator and allows a comparison with other metrics. On the other hand, the gamut-mapping optimization employing iCID as an objective function is compared with a state-of-the-art spatial gamut-mapping algorithm. Furthermore, the introduction of optimization intents for iCID-based gamut-mapping optimization allows the user to obtain results with desired color properties.

6.5.1 Properties of the iCID Metric

To summarize, elaborating the shortcomings of the CID-based gamut-mapping optimization (see Section 6.3.2) resulted in the *improved Color-Image-Difference* (iCID) metric (see Section 6.4). iCID is described by Equation 6.11 with the corresponding parameters given in Table 6.1.

However, the iCID metric used as an objective function for optimizing gamut mapping differs from the iCID metric used to predict perceived image differences. On the one hand, the goal of iCID-based gamut-mapping optimization is a pixel-wise optimal reproduction with respect to the metric independent of the viewing conditions. Therefore, image-appearance modeling must not be applied in the normalization step (see Section 6.3.1). On the other hand, iCID used as an IDM should have image-appearance modeling because viewing conditions play an important role (see Section 5.5) and small deviations at pixel scale might not have an impact on the perceived image difference.

Since there is missing information about viewing conditions in most of the databases, we considered only the visual resolution for image-appearance modeling if iCID is used as an IDM. For the achromatic and chromatic CSFs, a visual resolution $VR = 40$ samples per degree of visual angle was assumed. For the investigation of iCID as a metric, we refer to the CSFs proposed by the *iCAM* framework in the intensity-linear YCC opponent color space [117] – the same normalization was used for CID-CSF2 (see Section 5.4.2). As it is the only CSF filtering employed by iCID, the metric is denoted as *iCID-CSF*. For the investigation of the normalization's influence on the prediction performance of iCID, the results are also presented without CSF filtering called *iCID-None*.

Table 6.2: Hit Rates $\hat{p}^+$ of CID, iCID, and SSIM on the Combined Gamut-Mapping Database [112].

	CID	iCID*	iCID	SSIM	Majority Hit Rate
None	0.674	0.677	0.668	0.665	0.800
CSF	0.676	0.676	0.673		

*: without chroma-contrast and chroma-structure comparison

$^+$: hit-rate difference of 0.008 and larger is significant

6.5.2 iCID's Performance on Gamut-Mapping Distortions

Contrary to the investigation of CID, we used the whole Combined Gamut-Mapping Database (see Section 4.5) to test the performance of iCID. A separation into training and test data is not necessary because iCID's parameters were determined visually instead of being trained on a database. Hence, the hit rates (defined in Section 4.2.2) given in the following deviate from the hit rates computed for CID in Section 5.5. Also the majority hit rate changes to $\hat{p}_m = 0.800$[1]. According to Yule's two-sample binomial test (see Section 4.2.4) and assuming a significance level $\gamma = 0.05$, a predicted hit-rate difference of 0.008 and higher is significant and therefore not the result of chance.

The results of the investigation are shown in Table 6.2. The most important finding – and quite striking – is that CID and iCID do not perform significantly different (assuming the same normalization). I want to emphasize that CID was partly trained on the gamut-mapping data while iCID's parameters were adjusted independently.

The impact of the chroma-contrast and the chroma-structure comparison is significantly negative if no CSF filtering is performed (see first row of Table 6.2). We concluded that CSF filtering is crucial for evaluating gamut-mapping distortions and that is essential to maximize iCID's prediction performance. Otherwise, the influence of the chroma-contrast and the chroma-structure comparison – compared to the corresponding lightness analogs – is too high on the biased gamut-mapping database which results in a significantly negative impact.

Additionally, the SSIM index with a hit rate $\hat{p} = 0.665$ is included in Table 6.2 for comparison. It is the by far best performing IDM among all the other metrics – a detailed investigation on the Combined Gamut-Mapping Test Set can be found in Section 5.5.4. Even though SSIM is significantly worse than both the CID metric and the default iCID metric (with CSF filtering), its accuracy is astonishingly high on color distortions keeping in mind that SSIM totally disregards chromatic information. The design of the gamut-mapping database is a plausible explanation: the images are mapped to the same color gamut and almost all gamut-mapping algorithms preserve hue. Hence, the impact of chroma variations in the visual experiments was rather small compared to lightness-contrast and lightness-structure features. And

[1]This value is not exactly the same as in the original paper [112]. I detected the same pairs of distorted images in different datasets which were treated as different pairs. Thus, the majority hit rate is smaller.

these image-difference features are well predicted by the SSIM index. To sum up, chromatic features play a significant role – compare hit rates of iCID and SSIM – but not a dominant role – compare hit-rate ratios (defined in Section 4.2.2) of iCID (e.g., $\hat{p}_r(\text{iCID}-\text{CSF}) = 0.58$) and SSIM ($\hat{p}_r(\text{SSIM}) = 0.55$) – for assessing perceived image differences of common gamut-mapping distortions.

A multi-scale approach was neither investigated for the iCID metric nor employed in the iCID-based gamut-mapping optimization. Due to the findings that gamut-mapping distortions highly correlate across scales (see Section 5.5.2), the multi-scale approach may even adversely affect iCID's performance on the Combined Gamut-Mapping Database. Computing the hit rates for iCID with automatic downsampling like in Sections 5.5.3 also did not show significant difference to the hit rates computed without downsampling.

The majority hit rate did not change significantly. It is still far above the iCID metric which is performing best among all known IDMs. The influence of semantics and heuristics present in gamut-mapping databases on IDMs using only low-level features for predicting overall image quality is still an open research field.

6.5.3 iCID's Performance on Conventional Distortions

On conventional distortions such as blur, noise, and compression artifacts, some renowned IDMs show very high correlations to subjective data [129]. These metrics even do not consider chromatic information so I conclude that chromatic information is not essential. The chroma-difference and the hue-difference image-difference feature of the CID metric do not alter the prediction performance on conventional distortions (see Section 5.5.3) – except the multi-scale CID-None. But do the chroma-contrast comparison and the chroma-structure comparison of the iCID metric impair the accuracy on conventional distortions? We investigated this aspect on the recently introduced TID2013 (see Section 4.3.2) by computing iCID with and without chroma-contrast comparison and the chroma-structure comparison.

Section 5.5.2 revealed that image-difference features have a lesser correlation across scales for conventional distortions than for gamut-mapping distortions. Since a multi-scale version of iCID was not considered, we investigated at least an automatic downsampling to a scale the HVS is particularly sensitive to. We employed an automatic downsampling as in a recent implementation of the SSIM index [145]. As already mentioned, automatic downsampling has no significant influence on the hit rate for gamut-mapping distortions.

Table 6.3 summarizes the Spearman correlation on TID2013 also including SSIM and MSSIM for comparison as well as FSIMc [155] as the up to now best performing IDM [107]. If CSF filtering is not performed (first two rows), the additional chroma-contrast and chroma-structure comparison (second column) have a negative effect on the prediction performance compared to iCID without these comparisons (third column). The influence of both comparisons is obviously too high and thus impair the accuracy

Table 6.3: Spearman Correlations on the TID2013 [112].

	CID	iCID*	iCID	SSIM	MSSIM	FSIMc
None	0.604	0.656	0.501	0.627	0.785	0.851
None$^+$	0.752	0.813	0.705	0.742		
CSF	0.666	0.713	0.693			
CSF$^+$	0.773	0.823	0.813			

*: without chroma-contrast and chroma-structure comparison
$^+$: automatic downsampling

on conventional distortions. If the achromatic and chromatic components are blurred by CSF filtering (third and forth row) the influence is lower on conventional distortions and the prediction performance increases – contrary to gamut-mapping distortions (see Table 6.2). The slightly lower correlations of the additional chroma-contrast and chroma-structure comparisons which should not have an impact on conventional distortions suggest a still insufficient chromatic blurring. An improved performance of the iCID metric may be yielded by a more appropriate samples per degree value of the visual resolution but no information in this regard is provided by the TID2013 database.

The modifications made for iCID seem to be beneficial on conventional distortions compared to CID (first column) – assuming the correct treatment of chroma-contrast and chroma-structure comparison (omitting them or appropriate blurring) – while no significant difference exists on gamut-mapping distortions. This reveals the advantage of iCID's more general parameter determination by iCID-based gamut-mapping optimization rather than CID's parameter training on gamut-mapping data. The downsampled iCID with CSF filtering (Spearman correlation of 0.813) exceeds the performance of CID, SSIM, and MSSIM. Only the FSIMc index achieves a higher correlation to human judgments (0.851). However, the parameters of FSIMc were trained on a subset of TID2008 – hence a subset of TID2013 – which does not allow for a fair comparison. Furthermore, FSIMc has by far a worse prediction performance on the Combined Gamut-Mapping Database. Hence, I believe that iCID has the potential to be considered as a universal well-predicting IDM.

The use of automatic downsampling results in a great asset of iCID's correlation with human perception. Together with a suitable normalization of the input images' chromatic and achromatic channels according to the visual resolution the prerequisite for a high prediction performance is met. A convenient multi-scale approach may further increase the prediction performance on conventional distortions, e.g., MSSIM exceeds the performance of SSIM.

Scatter plots are more detailed and pictorial expressions of the metrics' behavior on subjective data. For the TID2013, scatter plots of MOS versus predictions are illustrated in Figure 6.7 including the metrics from Table 6.3 (CSF filtered and downsampled when available). Additionally, a fit with a logistic function as suggested in [129] is shown for easier comparison. The performance of an IDM is the better,

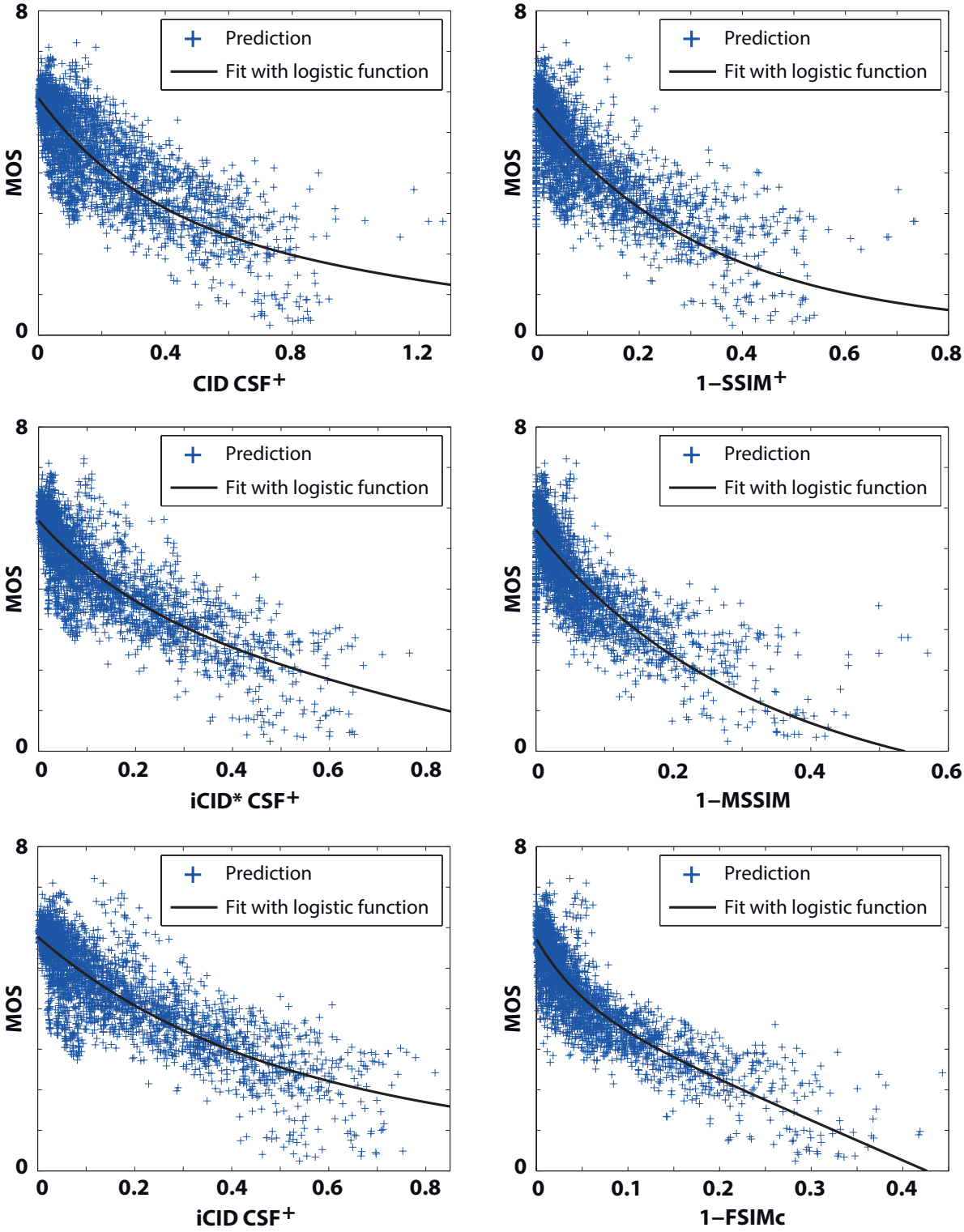


Figure 6.7: Scatter plots of subjective scores (MOS) against predictions of the investigated image-difference metrics on the TID2013. Each plot was fitted with a logistic function as suggested in [129]. The original version of the figure has already been published in [112].

the narrower the point cloud in the metric's plot. Visually, the most spread point clouds are engendered by CID-CSF$^+$ and SSIM$^+$ that is also reflected by the Spearman correlation in the corresponding table. For the plots of CID and iCID, there are outliers in the sector of MOS < 4 and (i)CID < 0.1. These predictions belong to the *local block-wise distortions of different intensity*, i.e., CID and iCID are not capable to properly assess this kind of distortion. The *change of color saturation* distortion is not detected by SSIM$^+$ and MSSIM indicated by the line of points situated very closely to the y-axis. A quite plausible result since both metrics totally disregard chromatic information – in opposite to CID, iCID, and FSIMc.

6.5.4 Results of iCID-Based Gamut-Mapping Optimization

Although the iCID metric shows no significant improvement on the Combined Gamut-Mapping Database and a moderate improvement on the TID2013, it offers a major advantage in comparison to CID: using iCID as an objective function for optimizing gamut mapping results in images without lightness-inversion, chromatic-ringing, chromatic-edge, and lightness-banding artifacts. Furthermore, the iCID-base gamut-mapping optimization greatly retains local contrast, structure, and color of the original image. These findings are more evident when the optimization is applied on a very small color gamut. In the following, we refer to such a color gamut specified by the `USNewsprintSNAP2007.icc` profile which covers less than 30 % of the sRGB gamut in the almost perceptually uniform LAB2000HL color space.

An example of iCID-based gamut-mapping optimization is given in Figure 6.8. Reference images are shown on the left, which were gamut mapped in a first step (middle column). The corresponding gamut-mapping transformation is incorporated in the ICC profile creating gamut-mapped images pixel-by-pixel with perceptual rendering intent. Then, the iCID-based gamut-mapping optimization was performed (right column) utilizing the gamut-mapped images as starting images. Obviously, the gray haze generated by the ICC profile almost disappears at the optimized images. Moreover, iCID-based gamut-mapping optimization produces images that retain chroma much more than the starting images, e.g., the color of the sky and some patches of the balloons. Fully exploiting the small color gamut is possible for the optimization because hue shifts are allowed for the benefit of gaining chroma – contrary to common gamut-mapping algorithms which include hue-preserving constraints.

A state-of-the-art spatial gamut-mapping algorithm by Zolliker and Simon named *Local Contrast Recovery* (LCR) [158] serves as a benchmark. LCR uses an unsharp masking technique based on edge-preserving smoothing which avoids formations of artifacts and needs an initial gamut-mapping transformation as well. iCID-based optimization was compared with LCR by applying the method on the same ICC profile used before. Figure 6.9 shows two exemplary comparisons including the reference images and both approaches. Contrast and structure are restored to a great extent and haze induced by ICC-profile gamut-mapping transformations is drastically reduced by both approaches. However, LCR has limited

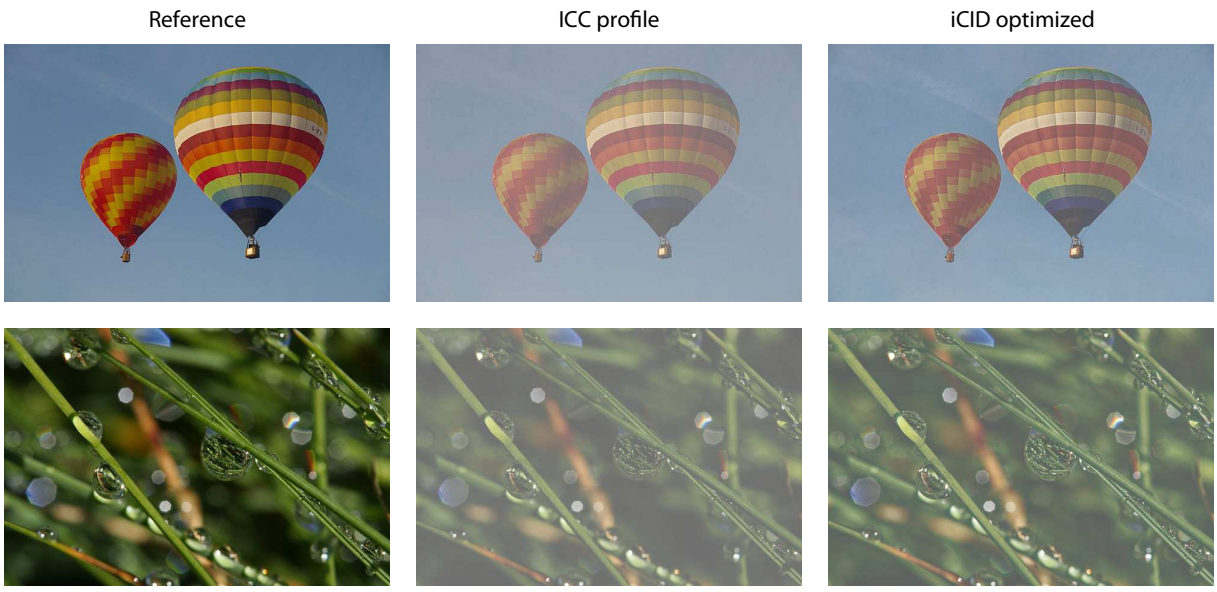


Figure 6.8: Examples of iCID-based gamut-mapping optimization (iCID optimized) compared with its starting images created by ICC-profile gamut-mapping transformations with perceptual rendering intent (ICC profile). Reference images were taken from Wikimedia Commons [11, 83]. Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The figure has originally been published in [112].

abilities to retain the colors of the reference images because it depends on the colors of the starting images which are rather dull. Local Contrast Recovery does not fully exploit the whole color gamut like iCID-based optimization is able to. Particularly, the black point of iCID-based gamut-mapping optimization is darker than for LCR as can be seen in the dark regions of the motorcycle image. The yellow parts of the motorcycle as well the sand and the cloth in the desert image are exemplary for iCID-based optimization's color-restoring ability.

To verify the performance of the proposed iCID-based gamut-mapping optimization, a visual experiment was conducted on 14 natural images showing various scenes. For that purpose, the ICC-profile transformation with perceptual rendering intent (denoted as ICC profile), Local Contrast Recovery (denoted as LCR), and iCID-based gamut-mapping optimization (denoted as iCID optimized) were computed for all images on the small color gamut specified by `USNewsprintSNAP2007.icc`. Please note that the ICC profile serves as the initial gamut mapping for both LCR and iCID optimized. Employing the methodology of paired comparisons (see Section 4.1), a reference and two corresponding distorted images were shown simultaneously on a calibrated LCD monitor. 15 color-normal unbiased observers were asked to choose the distorted image which is perceived more similar to the reference. In case of no preference the subject selected *tie* which counted half a vote for both methods.

For the analysis of the experiment, the evaluations are collected in a so-called *frequency matrix* which counts the number of decisions one method was preferred over another. Compared with ICC profile, the

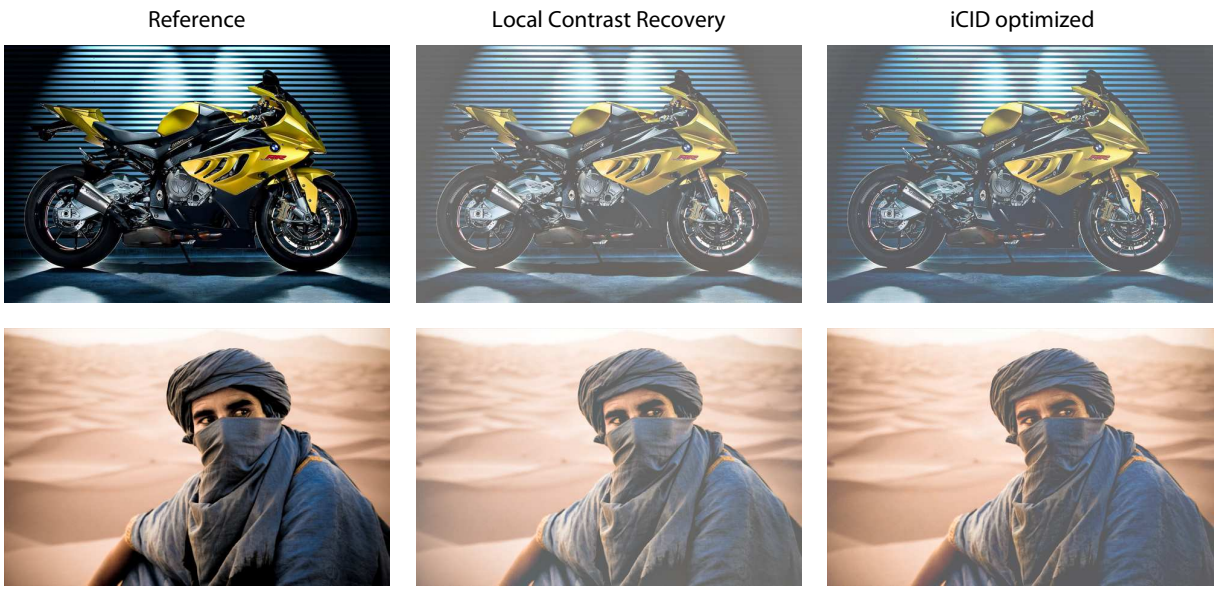


Figure 6.9: Examples of iCID-based gamut-mapping optimization (iCID optimized) compared with Local Contrast Recovery exploiting the same ICC-profile gamut-mapped starting image. Reference images were taken from Wikimedia Commons [136] and from Fotopedia [133]. Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The figure has originally been published in [112].

observers selected iCID optimized in 95 % of the cases and compared with LCR in 72 % of the cases. In the next step, so-called *accuracy scores* – averaged z-scores – are calculated based on the frequency matrix as described by Morovič [87]. The procedure of calculating accuracy scores has the advantage that it allows the determination of reliable confidence intervals. The results which are illustrated in Figure 6.10 reveal that iCID-based gamut-mapping optimization is evaluated significantly better than ICC profile and the state-of-the-art LCR.

The performance of optimizing gamut mapping by using the iCID metric as a target function was validated not only on the `USNewsprintSNAP2007.icc` profile but also on the slightly larger FOGRA27 ICC profile. In the upcoming chapter, iCID-based gamut-mapping optimization is further investigated by extending the optimization to high-dynamic-range imaging.

6.5.5 Optimization Intents

Contrary to most gamut-mapping algorithms, iCID-based gamut-mapping optimization allows hue shifts for the benefit of improved chroma or contrast. Thereby, the predicted image difference to the reference image is minimized employing iCID's image-difference features. However, these low-level features do not consider the images' semantics which may have a negative impact on perceived image-differences [139]. Semantics include, among others, memory colors of objects such as skin or brand colors. If, for instance, skin turns greenish due to iCID-based optimization, the image difference might be perceived

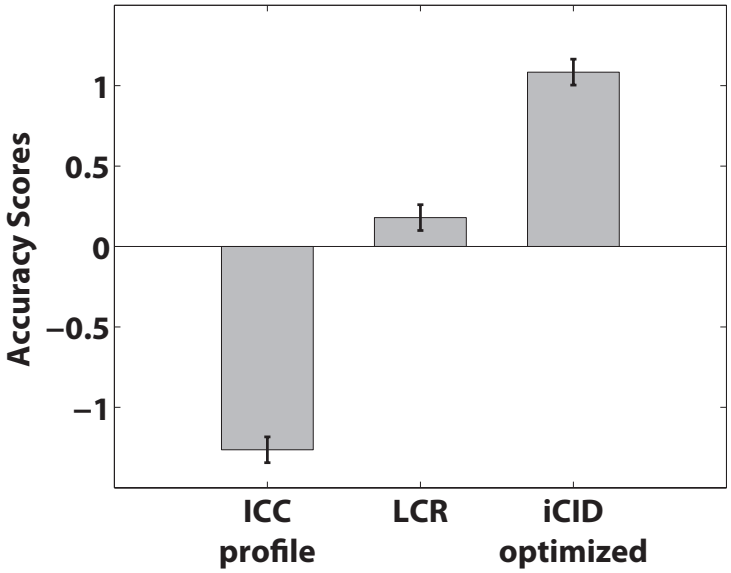


Figure 6.10: Bar plot of the accuracy scores gained from the visual paired-comparison experiment including ICC-profile gamut-mapped images (ICC profile), Local Contrast Recovery (LCR), and iCID-based gamut-mapping optimization (iCID optimized). The 95 % confidence interval (corresponds to a significance level $\gamma = 0.05$) is represented by error bars. The original version of the figure has already been published in [112].

very disturbing even though contrast and structure are improved.

Thus, I have suggested to control the strength of allowed chroma and hue shifts. This is possible by varying iCID's chroma-difference and hue-difference parameters (c_4 and c_5). The importance of chroma and hue is increased by larger parameter c_4 and c_5, respectively, leading to a stronger preservation of the color attribute. Three parameter sets of c_4 and c_5 were proposed while all other parameters remain unaltered. They are presented in Table 6.4 and denoted as *optimization intents* for iCID-based gamut-mapping optimization:

- **Perceptual:** equal weighting of lightness, chroma, and hue. Refers to the default parameters given in Table 6.1 within the perceptually uniform LAB2000HL color space.
- **Hue-preserving:** inhibits hue deviations.
- **Chromatic:** enforces smaller chroma and hue differences to the reference image leading to more chromatic optimization results. Hence, hue shifts are hardly admitted at the expense of larger variations in lightness.

How the optimization intents change the results of iCID-based optimization is illustrated in Figure 6.11. A further improvement of the proposed method might be made by employing semantic maps of the image such as skin or face detection algorithms. Based on these maps, spatially varying parameter

Table 6.4: Optimization Intents and Corresponding Parameters for iCID-Based Gamut-Mapping Optimization [112].

Optimization intent	c_4	c_5
perceptual	0.002	0.002
hue-preserving	0.002	0.02
chromatic	0.02	0.02

sets could be applied to preserve memory colors where necessary and allow hue shifts elsewhere for better results.

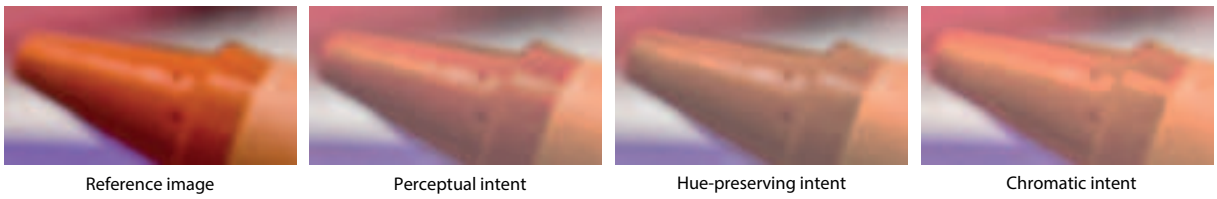


Figure 6.11: Example of the proposed optimization intents: perceptual, hue-preserving, and chromatic. The crayon's orange tip changes in chroma and hue according to the optimization intents. The image was taken from Fotopedia [60]. Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The original version of the figure has already been published in [112].

6.6 Conclusions

The good prediction performance on gamut-mapping distortions of the previously introduced *Color-Image-Difference* (CID) metric (see Chapter 5) made it interesting for optimizing gamut mapping. Hence, CID was used as an objective function to minimize the perceived difference to the original image subject to in-gamut images. For that purpose, an iterative algorithm was presented starting with an image gamut-mapped by a common gamut-mapping transformation which is optimized by employing a local step length below the just-noticeable color difference. A descent direction of the global CID metric is guaranteed even though the metric is updated only on a small subimage after altering one pixel.

Optimized images were judged to be more similar to the original than the starting images. Nonetheless, the optimization created disturbing artifacts such as lightness banding and chromatic ringing. Addressing these artifacts suggested to introduce multiple modifications to CID resulting in the *improved Color-Image-Difference* (iCID) metric. The parameters were adjusted so that the *iCID-based gamut-mapping optimization* yielded artifact-free images. Thus, the iCID metric is not biased by heuristics included in image-quality-assessment databases like it would be the case for the common methodology of parameter training.

Even though the parameters were not trained, the prediction performance on gamut-mapping distortions as well as on conventional distortions was increased by the modifications. Please note that an appropriate chromatic and achromatic normalization of the input images is needed to account for the visual resolution. Since iCID consists of two additional chromatic image-difference comparisons, its predictions accuracy is more sensitive to the viewing conditions than CID.

The capability of iCID-based gamut-mapping optimization is to obtain artifact-free in-gamut images that restore contrast, structure, and especially color of the starting images to a great extent. In a visual experiment, iCID-based optimization was compared to a common gamut-mapping transformation and a state-of-the-art spatial gamut-mapping algorithm with respect to perceived similarity to the original image. 15 observers verified that iCID-based gamut-mapping optimization significantly outperforms the other gamut-mapping algorithms. In addition, I presented optimization intents which allow to choose a preferred color configuration for the iCID-based gamut-mapping optimization.

In my opinion, the unconventional methodology of using an image-difference metric as an objective function for gamut-mapping optimization helps to learn more about the underlying image-difference metric as well as the importance of single (low-level) color-reproduction attributes. Moreover, the parameters adjusted to yield the optimal reproduction allow a more general parameter determination of the image-difference metric not biased by image-quality-assessment databases. A further improvement of the iCID-based optimization may be achieved by employing semantic information of the image. For instance, using face detection and pattern recognition in combination with spatially-varying iCID parameters which adjust the desired optimization intent (e.g., more weight is given to the hue-difference comparison for memory colors). Future research should focus on higher-level features for predicting image differences and for optimizing gamut mapping.

6.7 My Contributions

To emphasize my contributions to the papers referring to this chapter (see Section 6.1), they are summarized in the following:

- Programming of the C++ implementation for (i)CID-based gamut-mapping optimization.
- Solving the chromatic-ringing, chromatic-edges, and lightness-banding artifacts.
- Final parameter adjustment for the iCID metric.
- Evaluation of iCID's performance on conventional and gamut-mapping databases.
- Quantitative analysis resulting in accuracy scores and scatter plots.
- Introduction of optimization intents.

Furthermore, these are my investigations which I presented in this chapter and which have not been published yet:

- Recomputation of the majority hit rate.
- More detailed explanation of efficiently programming the gamut-mapping optimization.
- Computation of hit-rate ratios.

7

High-Dynamic-Range Gamut Mapping

In the previous chapter, gamut mapping has been proposed as a constrained optimization problem corresponding to Equation 6.2. Using the *improved Color-Image-Difference* (iCID) metric as the objective function, a visual experiment revealed that the optimized gamut-mapped images were judged to be perceptually closer to the reference images than the starting image and even a state-of-the-art gamut-mapping algorithm. This chapter presents the application of iCID-based gamut-mapping optimization for high-dynamic-range images. First, the standard reproduction workflow of high-dynamic-range images is presented which leaves room for improvement. Then, perceptual high-dynamic-range color spaces are presented which allow a simultaneous representation of high-dynamic-range and low-dynamic-range images. Thereby, a new concept of tone mapping is proposed as gamut mapping in a high-dynamic-range color space – called *high-dynamic-range gamut mapping*. To verify the new concept, iCID-based gamut-mapping optimization is extended to high-dynamic-range imaging and tested in a visual experiment on a high-dynamic-range display. The results of the experiment help to further improve the iCID-based high-dynamic-range gamut-mapping optimization.

7.1 Related Publication

Most of the work introduced in this chapter refers to the conference paper by Preiss et al. [111]:

> J. Preiss, M. D. Fairchild, J. A. Ferwerda, and P. Urban. Gamut Mapping in a High-Dynamic-Range Color Space. In *Proc. SPIE 9015, Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications*, p. 90150A, 2014.

This chapter is written in the same way as the previous one: most investigations refer to already published content in the given paper. New content as well as my main contributions are particularly emphasized.

Please note that the conference paper introduces a new concept with an application rather than a thorough analysis of the topic. Hence, this chapter does not claim to contain a complete investigation of the topic.

7.2 Standard Reproduction Workflow of High-Dynamic-Range Images

Real-world scenes can exhibit an enormous range of luminance levels, e.g., from direct sunlight in the sky to a dark object in the shadow. The ratio between the brightest and the darkest point is called *dynamic range*. However, most devices (e.g., standard cameras, printers, or standard displays) are only able to capture or reproduce a few magnitudes of dynamic range – which is referred to as *low-dynamic-range* (LDR) imaging. With advancing imaging technology, it is nowadays possible to capture, store, or reproduce images which represent a greater range of luminance levels – which is referred to as *high-dynamic-range* (HDR) imaging.

HDR images representing real-world scenes are not reproducible without any error by LDR devices due to the dynamic range of the HDR images and colors that possibly exceed the devices' color gamut. Therefore, HDR images have to be distorted to fit into the reproducible dynamic range and color gamut of the LDR device. The standard workflow of such a distortion usually consists of two subsequent transformations: 1. *HDR tone mapping*, and 2. *color gamut mapping* (see Section 2.4) [111]. The first step is applied by so-called *tone-mapping operators* (TMOs) which strongly reduce the images' contrast to map the HDR into the LDR and the second step is applied by gamut-mapping algorithms (GMAs). The magenta box in Figure 7.1 illustrates this standard workflow. However, using these subsequent mappings may not be optimal for representing HDR images in the output devices' LDR color gamut. In particular, most TMOs only map the luminance channel while disregarding the chromatic channels, and the GMAs do not have access to the color contrast ratios within the HDR images as they only operate on already tone-mapped images.

A detailed survey of already proposed TMOs is given by Reinhard et al. [118] – only a small number of them accounts for color. Neglecting chromatic information may lead to visually disturbing color shifts of the tone-mapped images due to luminance-induced appearance phenomena [111]. This shortcoming is usually solved by applying color correction [82,116,123] or by using color-appearance or image-appearance models for tone mapping [2, 66, 97]. Nonetheless, the applicability of these color-related approaches is questionable since color-corrected tone-mapped images are rather visually pleasant than perceptually accurate and numerous viewing-condition parameters are required for color-appearance and image-appearance models.

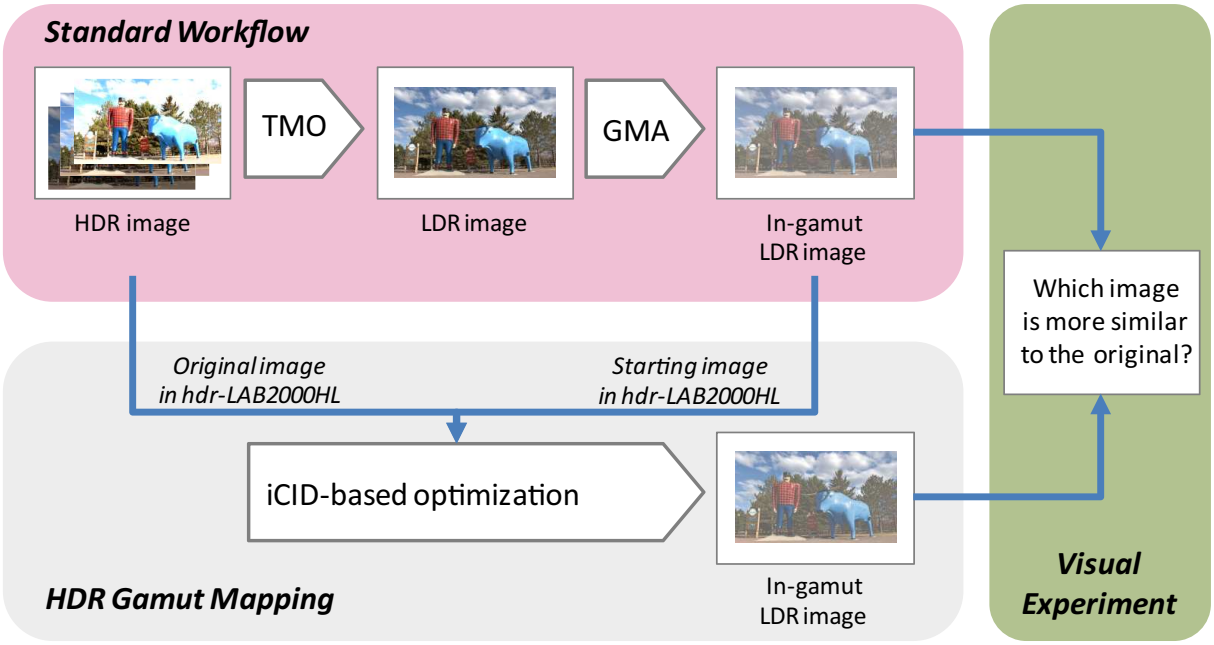


Figure 7.1: 1. Illustration of the standard workflow to represent a high-dynamic-range (HDR) image in the color gamut of an output device (magenta box). First, a tone-mapping operator (TMO) is applied. Then, a gamut-mapping algorithm (GMA) is applied. 2. Proposed HDR gamut mapping by applying iCID-based HDR gamut-mapping optimization (gray box). 3. Visual experiment to compare the results of the standard workflow with the results form the iCID-based HDR gamut-mapping optimization (green box). The HDR image was taken from the *HDR Photographic Survey* proposed by Fairchild [39]. The original version of the figure has already been published in [111].

The ability of GMAs to retain color contrast ratios of the reference image is limited if they transform the tone-mapped LDR image instead of the reference HDR image. Especially the aforementioned inappropriate handling of color by the employed TMO may result in a misinterpretation of perceived color contrast ratios within the HDR scene by the GMA.

An easier handling of HDR-image reproduction can be achieved by a simultaneous representation of HDR and LDR images as well as of the devices' color gamut. Furthermore, this enables a unified transformation of HDR images into the output devices' color gamut. For this purpose, the concept of perceptual HDR color spaces – which allow a simultaneous representation with coordinates related to perception – is presented in the following section.

7.3 Perceptual High-Dynamic-Range Color Spaces

Perceptual color spaces – i.e., color spaces with coordinates linked to intuitive color attributes – were mainly designed for LDR applications, such as CIELAB or LAB2000HL (see Section 2.3). Even if it is possible to deal with intensities brighter than diffuse white – e.g., allowing lightness values $L^* > 100$

in CIELAB – these color spaces are likely to fail in HDR applications due to misinterpretation of HDR stimuli [41]. For the emerging field of HDR imaging, however, appropriate perceptual HDR color spaces could be beneficial for processing HDR images [118]. Thus, some perceptual HDR color spaces are presented in the following.

Mantiuk et al. [81] developed a perceptual HDR color space in 2007 that extends existing image and video compression standards such as *JPEG-2000* or *MPEG-4*. The luminance is converted into *luma* (see Section A.2.1 in the appendix) – as a 12-bit integer code value – by a transfer function derived from contrast detection models. The chromatic channels are encoded 8-bit employing the approximately perceptually uniform *CIE 1976 Uniform Chromaticity Scales* (UCS) $u'v'$.

A concept of a perceptual HDR color space was recently introduced by Fairchild and Wyble [41]. They extended the CIELAB color space (see Section 2.3.3) by replacing the CIELAB cube-root-based function $f(\xi)$ (see Equation A.15 in the appendix) by an appropriately parametrized Michaelis-Menten function [92]. In 2011, Fairchild and Chen [40] presented an improved version of this HDR extension of CIELAB denoted as *hdr-CIELAB* which is almost similar to CIELAB when operating in the LDR space. In addition, they presented an analogous HDR extension of the *IPT* color space [32] denoted as *hdr-IPT*.

For the purpose of creating a perceptually uniform and hue linear perceptual HDR color space, I adapted the concept of hdr-CIELAB to the LAB2000HL color space (presented in Section 2.3.4). LAB2000HL was designed to improve CIELAB with respect to perceptual uniformity and hue linearity. The transformation into the resulting perceptual HDR color space called *hdr-LAB2000HL* [111] is given in Appendix A.2.5.

Even though hdr-LAB2000HL is based on a hue linear and almost perceptually uniform color space, it does not have to possess these properties itself. In particular, these properties have not been proved for a lightness larger than diffuse white. The hdr-LAB2000HL perceptual HDR color space is tested by using it for the application presented in the upcoming section.

7.4 Gamut Mapping in a Perceptual High-Dynamic-Range Color Space

To overcome the shortcomings of a two-step strategy for the reproduction of HDR images, we have suggested to perform tone mapping and then gamut mapping in a single step. This approach denoted as *HDR gamut mapping* requires a simultaneous representation of HDR and LDR images as well as of the devices' color gamut, in particular a perceptual high-dynamic-range color space. HDR gamut mapping can be performed on HDR images analogously to color gamut mapping performed on LDR images (see Section 2.4) – but not many investigations of gamut mapping in the scope of HDR have been made so far.

To test if color contrast ratios of the HDR image can be retained by HDR gamut mapping, we extended an existing GMA to HDR. Moreover, we performed the HDR gamut mapping in the presented hdr-LAB2000HL to verify its suitability.

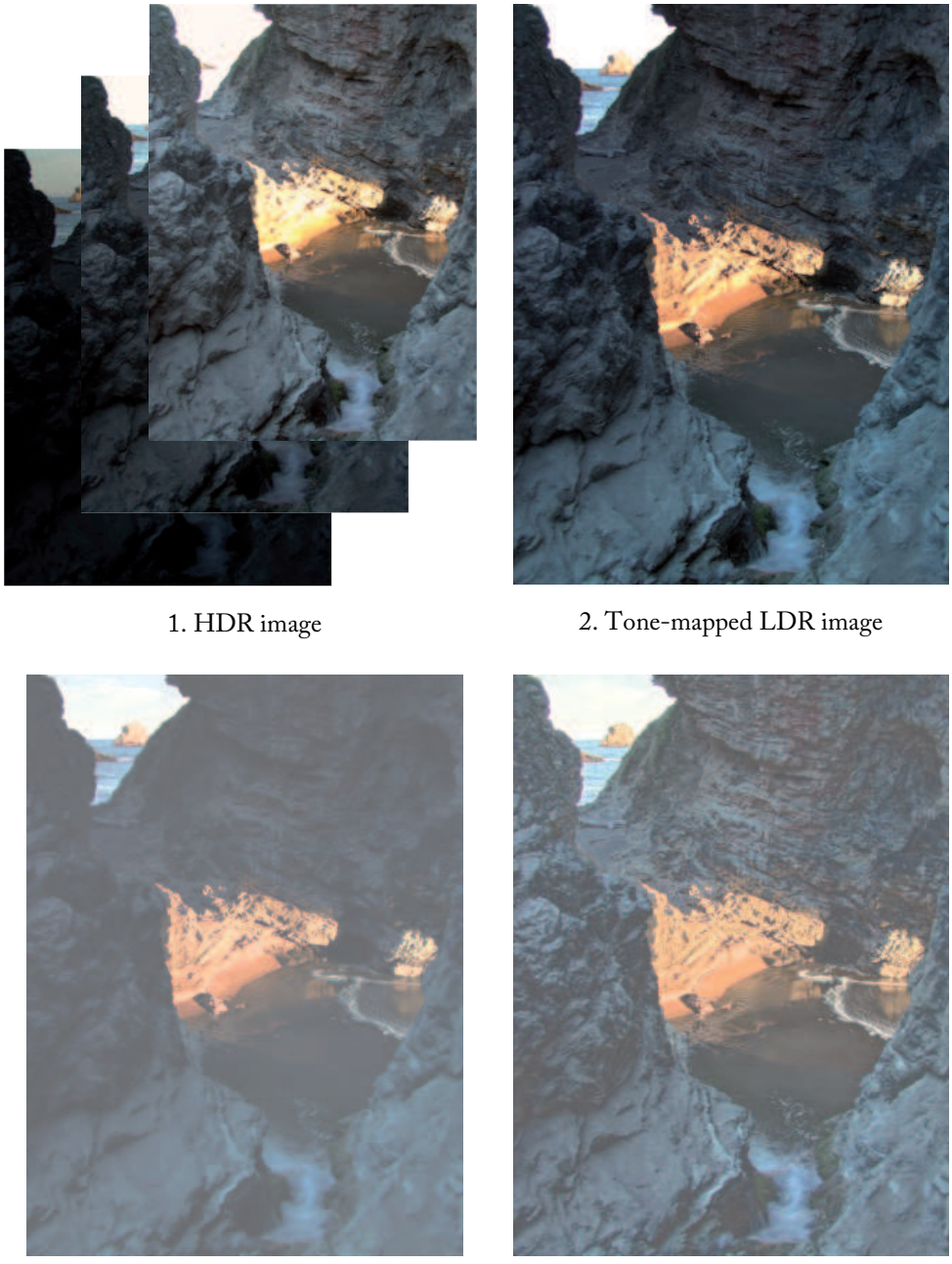
In particular, we adapted the iCID-based gamut-mapping optimization (introduced in Chapter 6) to HDR images – hence, denoted as *iCID-based HDR gamut-mapping optimization*. The constrained optimization problem (defined in Equation 6.2) applied to iCID and extended to the HDR now reads:

$$\mathcal{Z} = \underset{\mathbf{Y}\widetilde{\subset}\mathcal{G}}{\operatorname{argmin}}\ \mathrm{iCID}(\mathcal{X}_{\mathrm{HDR}}, \mathbf{Y}) \tag{7.1}$$

where $\mathcal{X}_{\mathrm{HDR}}$ is the reference HDR image, $\mathcal{Z}$ is the optimized in-gamut LDR image, the expression $\mathbf{Y}\widetilde{\subset}\mathcal{G}$ indicates that all colors of the LDR image $\mathbf{Y}$ are within the color gamut $\mathcal{G}$ of the output device, all images as well as the color gamut are represented in the hdr-LAB2000HL color space, and all images have the same image resolution. Thus, assuming an appropriate performance of iCID on the HDR, the hdr-LAB2000HL color space is proved to be suitable for HDR gamut mapping if the iCID-based optimized in-gamut LDR image is perceived closer to the reference HDR image than the starting LDR image.

For our investigation of HDR gamut mapping and hdr-LAB2000HL, the following settings for the iCID-based HDR gamut-mapping optimization (see gray box in Figure 7.1) have been proposed. Before transforming the HDR images into the hdr-LAB2000HL color space, a global luminance adaptation was applied: the geometric mean of the image pixels' luminance in the CIEXYZ color space was constrained to equal the luminance of the middle gray value of hdr-LAB2000HL [111]. A chromatic-adaptation transform to account for different scene illuminants was not considered. The starting image of the iterative optimization was obtained by first applying a standard TMO to the original HDR image using the *HDR Toolbox* from Banterle [7] and then applying a standard GMA to the tone-mapped LDR image. The iterative optimization was preset to terminate after reaching a number of 20 iterations. The steps of the proposed HDR gamut mapping are shown in Figure 7.2 for an exemplary HDR image.

The concept of HDR gamut mapping as well as the hdr-LAB2000HL color space are employed by the proposed iCID-based HDR gamut-mapping optimization. To verify concept and color space, the results of the optimization were compared with their corresponding starting images in a visual experiment as described in the upcoming section.



1. HDR image

2. Tone-mapped LDR image

3. Tone- and gamut-mapped LDR image

4. iCID-based optimized in-gamut LDR image

Figure 7.2: Steps of the proposed high-dynamic-range (HDR) gamut mapping to transform an HDR image into the color gamut of a low-dynamic-range (LDR) device: 1. HDR reference image; 2. tone mapping; 3. gamut mapping; 4. iCID-based HDR gamut-mapping optimization by minimizing the perceived image difference to image 1 and starting with image 3. The HDR image was taken from the DVD-ROM included in the book *High Dynamic Range Imaging* by Reinhard et al. [118]. Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The original version of the figure has already been published in [111].

7.5 Visual Experiment

The proposed iCID-based HDR gamut-mapping optimization minimizes the iCID metric of in-gamut LDR images to the reference HDR image within the hdr-LAB2000HL color space. To test the performance of this optimization, the resulting optimized images were compared with the corresponding starting images – obtained by tone mapping and then gamut mapping of the reference HDR images – in a visual experiment (see green box in Figure 7.1). The observers had to decide which LDR reproduction is perceived more similar to the reference HDR image without allowing tie decisions. The images were shown on an HDR display [44] at the *Munsell Color Science Laboratory* in Rochester – the reference HDR image was placed in the middle of the display and the iCID-based optimized in-gamut LDR image as well as the corresponding starting LDR image were placed to the left and to the right of the HDR image, respectively. Both positions were used for each LDR representation but in random order.

For the realization of the visual experiment, I chose twelve reference HDR images which exhibit various natural scenes. Moreover, three different TMOs were applied to the reference HDR images and two different gamut-mapping transformations were then applied to the tone-mapped LDR images. The HDR images were taken from the *HDR Photographic Survey* introduced by Fairchild [39] as well as from the DVD-ROM included in the book *High Dynamic Range Imaging* by Reinhard et al. [118]. The TMOs include: 1. Reinhard's bilateral TMO, 2. Drago's TMO (both TMOs were adopted from the HDR Toolbox [7]), and 3. tone mapping provided by *iCAM06* [66]. The gamut-mapping transformations include: 1. a gamut-mapping transformation incorporated in the ICC profile `USNewsprintSNAP2007.icc` which specifies a very small newspaper gamut, and 2. the color-space transformation from CIEXYZ to sRGB (given in Appendix A.2.2) representing the mapping to a standard LDR monitor.

The visual experiment was conducted by 13 observers. Thereby, each observer was asked to make a total of 144 decisions (12 images $\times$ 2 reversed orders $\times$ 3 TMOs $\times$ 2 GMAs). The results of the experiment are presented in the subsequent section.

7.6 Results and Discussion

In the visual experiment, the iCID-based optimized in-gamut LDR images were compared with the corresponding starting LDR images – which represent the standard reproduction workflow (see Section 7.2) – in relation to the reference HDR images. For each of these pairs, the ratio of the number of choices favoring the optimized image to the number of all choices was computed. Then, the image pairs were categorized into groups of color gamuts (small newspaper gamut and sRGB gamut), TMOs (Reinhard's bilateral TMO, Drago's TMO, and tone mapping provided by iCAM06), and image scenes (dim and others) [111]. Table 7.1 shows the categorized results.

Table 7.1: Preference of the HDR Gamut-Mapping Optimization to the Standard HDR Workflow [111].

	Image Scenes	all	dim	others	all		
	TMOs	all	all		Reinhard	Drago	iCAM06
Color Gamuts	Newspaper	0.95	0.90	0.98	0.93	0.95	0.96
	sRGB	0.52	0.35	0.63	0.51	0.53	0.51

The results of the very small newspaper gamut which was chosen to better illustrate the differences of gamut-mapped images are shown in the first row. For any TMO and any group of image scenes, the vast majority of the observers preferred iCID-based optimization to its starting images. This finding reveals that HDR gamut mapping and the hdr-LAB2000HL color space are not only suitable for transforming HDR images into the output devices' color gamut but also able to significantly improve the standard workflow. The proposed iCID-based HDR gamut-mapping optimization seems to retain color contrast ratios as seen in the example of Figure 7.2.

However, the behavior is quite different for the sRGB device gamut (second row of Table 7.1). Including all TMOs and image scenes (first column), only 52 % of the observers favored the proposed HDR gamut mapping to the standard workflow. By regarding the iCID-based optimized in-gamut LDR images of the CIEXYZ to sRGB gamut-mapping transformation, a group of images resulted in artifacts such as halos and over-sharpening. This group of images exhibiting dim image scenes achieved only a preference of 35 % (second column) while the other images which do not possess artifacts achieved a preference of 63 % (third column) – see Figure 7.3 for a good example. I.e., even though the perceived difference between images from HDR gamut mapping and the standard workflow is small, HDR gamut mapping was indeed preferred except for dim image scenes.

A further investigation of the iCID-based optimized images possessing artifacts revealed that an inappropriate number of iterations was chosen for the images used in the visual experiment. Exemplary for the *Ahwahnee Great Lounge* image from the HDR Photographic Survey which exhibits a dim image scene, the iCID scores versus the number of iterations are shown in Figure 7.4 for the iCID-based HDR gamut-mapping optimization using Drago's TMO and the sRGB color gamut. As can clearly been seen, 20 iterations as used for the visual experiment are not sufficient even though the iCID scores are decreasing. For instance, an iteration number of 200 noticeably reduces the iCID scores while no further significant improvement is expected for even higher iteration numbers.

An improvement of the proposed iCID-based HDR gamut-mapping optimization may be considering chromatic adaptation. Thus, we also applied the CIECAM02 chromatic-adaptation transform exemplary to the Ahwahnee Great Lounge image (A to D65 chromatic adaptation due to the tungsten in-door lighting). Figure 7.5 shows the resulting LDR representations of the Ahwahnee Great Lounge image: 1. Standard reproduction workflow. 2. Proposed HDR gamut mapping which has an artificial looking

1. Tone- and gamut-mapped LDR image

2. iCID-based optimized in-gamut LDR image (20 iterations)

Figure 7.3: Comparison of 1. starting tone- and gamut-mapped image, and 2. proposed iCID-based optimized image. Color was improved by iCID-based HDR gamut-mapping optimization compared with the standard workflow, e.g., in the sky and in the clouds. Tone mapping was performed by the Drago tone-mapping operator and gamut mapping by the color-space transformation from CIEXYZ to sRGB. The HDR image (*Paul Bunyan*) was taken from the *HDR Photographic Survey* introduced by Fairchild [39]. Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. The figure has originally been published in [111].

due to halo and over-sharpening artifacts. 3. Proposed HDR gamut mapping with 200 iterations which drastically reduces the artifacts. 4. Proposed HDR gamut mapping with 200 iterations and chromatic adaptation which also reduces the hue shifts towards yellow.

The color contrast ratios in the dark as well as in the bright image regions are retained to a great extend in the third and forth image of Figure 7.5. This finding indicates that the hdr-LAB2000HL color space is suitable for the purpose of HDR gamut mapping [111]. However, artifacts are still apparent in the optimized images. Further improvements of iCID-based HDR gamut-mapping optimization may be achieved by applying local luminance adaptation [118] rather than global luminance adaptation, and by adjusting iCID to HDR applications, e.g., with a cut-off for ΔL in Equation 5.8 to not overestimate the lightness-difference image-difference feature.

Finally, the impact of different TMOs on the optimization result was analyzed represented by columns 4 - 6 in Table 7.1. It is obvious that the preference for iCID-based optimization is independent of the employed TMO because the ratios are quite similar for both color gamuts.

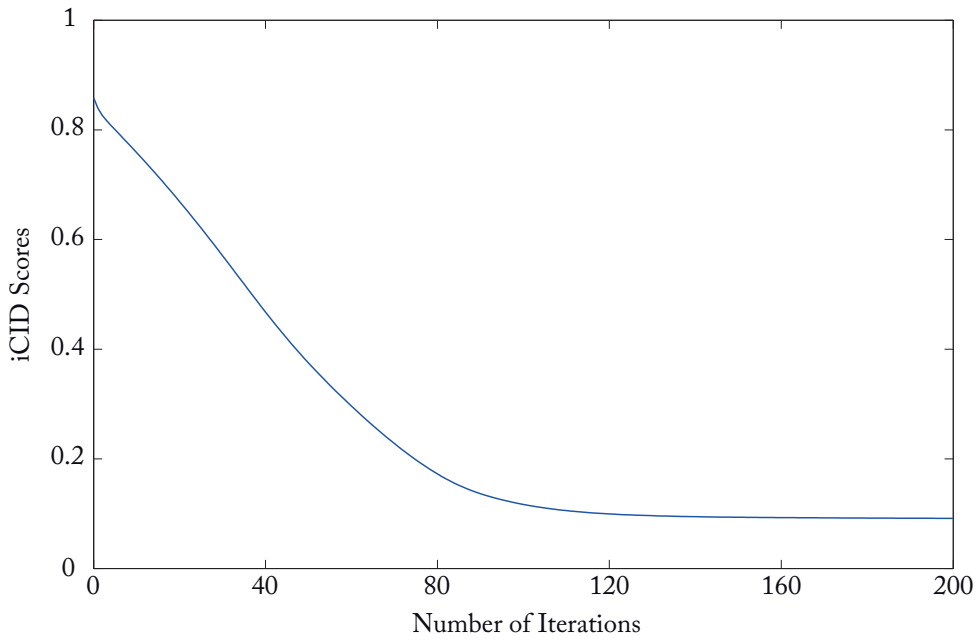


Figure 7.4: iCID scores versus the number of iterations for the iCID-based high-dynamic-range gamut-mapping optimization of the tone- and gamut-mapped *Ahwahnee Great Lounge* image with respect to its high-dynamic-range reference. Tone mapping was performed by the Drago tone-mapping operator and gamut mapping by the color-space transformation from CIEXYZ to sRGB. The original version of the figure has already been published in [111].

1. Tone- and gamut-mapped LDR image

2. iCID-based optimized in-gamut LDR image (20 iterations)

3. iCID-based optimized in-gamut LDR image (200 iterations)

4. iCID-based optimized in-gamut LDR image (200 iterations, chromatic adaptation)

Figure 7.5: Modifications of iCID-based high-dynamic-range (HDR) gamut-mapping optimization exemplary for the *Ahwahnee Great Lounge* image represented by in-gamut low-dynamic-range (LDR) images: 1. starting tone- and gamut-mapped image; 2. standard iCID-based optimized image; 3. iCID-based optimized image using more iterations; 4. iCID-based optimized image using more iterations and chromatic adaptation. Tone mapping was performed by the Drago tone-mapping operator and gamut mapping by the color-space transformation from CIEXYZ to sRGB. The HDR image was taken from the *HDR Photographic Survey* introduced by Fairchild [39]. Most accurate viewing conditions are guaranteed on a display calibrated to sRGB. Please note that the black pixels in the highlights are artifacts created by the color-space transformations and not by the optimization. The figure has originally been published in [111].

7.7 Conclusion

The standard workflow to transform a *high-dynamic-range* (HDR) image into the color gamut of a *low-dynamic-range* (LDR) device for the purpose of reproduction was presented. It pursues a two-step strategy of tone mapping and then gamut mapping which may lead to a misinterpretation of perceived color contrast ratios within the HDR image.

Hence, a new concept denoted as *HDR gamut mapping* has been proposed which performs a unified transformation within a so-called *HDR color space*. Moreover, a new perceptual HDR color space called *hdr-LAB2000HL* has been proposed which is an HDR extension of the almost perceptually uniform and hue linear LAB2000HL color space. To test both the new concept and the new HDR color space, iCID-based gamut-mapping optimization was adapted to HDR gamut mapping within the hdr-LAB2000HL color space denoted as *iCID-based HDR gamut-mapping optimization*.

A visual experiment on an HDR display was conducted to compare the results of this HDR gamut-mapping application with the standard workflow. The findings of the experiment indicated the suitability of hdr-LAB2000HL for HDR gamut mapping and revealed that HDR gamut mapping is able to retain perceived color contrast ratios of the reference HDR image.

An analysis of the shortcomings of some iCID-based optimized in-gamut LDR images, such as halo and over-sharpening artifacts, suggested some improvements like a larger iteration number for the proposed HDR gamut mapping. Future work shall consider an appropriate local luminance adaptation for the simultaneous representation of HDR and LDR images in the hdr-LAB2000HL color space. Furthermore, an HDR extension of iCID shall be investigated because there is a need of a reliable HDR-image quality assessment particularly regarding color [94].

I believe that HDR gamut mapping is a promising new topic in the emerging field of HDR imaging. In particular, HDR gamut mapping encoded within look-up tables like employed by industrial color management systems allows for faster processing and thus may serve as an HDR extension of ICC color management.

7.8 My Contributions

To emphasize my contributions to the paper referring to this chapter (see Section 7.1), they are summarized in the following:

- Adaptation of the almost perceptually uniform LAB2000HL color space to a perceptual high-dynamic-range color space denoted as hdr-LAB2000HL.
- Preparation and realization of the visual experiment on a high-dynamic-range display.

- Analysis of the experimental results.

8

Conclusion

This chapter reflects the main outcome of my work. First, the presented work is summarized and the results are discussed. Then, an outlook on future work and applications of image-difference metrics is given. Finally, I particularly emphasize the scientific added value of my work.

8.1 Summary

The aim of this work was to introduce a new *image-difference metric* (IDM) which is denoted as *improved Color-Image-Difference* (iCID) metric. The major advantage of iCID is that image normalization is applied and that chromatic information is employed instead of only evaluating the achromatic intensity. Moreover, iCID was used as an objective function for optimizing gamut mapping as well as *high-dynamic-range* (HDR) tone mapping. The optimized images were proved in several visual experiments to have a significantly lower perceived difference to the original image than state-of-the-art transformations.

To understand how iCID was derived, specific knowledge of visual psychophysics and color spaces was presented in Chapter 2. iCID is an IDM, i.e., an objective assessment of perceived differences of two images. In image processing, image difference corresponds to relative-to-reference image quality. An overview of image quality (see Section 3.1) and objective assessment of image quality (*image quality assessment* (IQA), see Section 3.1.1) was given. Chapter 3 also covers related work to IDMs and in particular IDMs assessing color information. However, objective IQA does not represent the ground truth of image quality – predictions of IQAs do not perfectly correlate with human image perception. To gain information about perceived image quality, subjective data has to be collected in visual experiments. How these experiments are conducted and how this subjective data is analyzed was described in Chapter 4. The results of such an experiment conducted by many subjects as well as all images involved in the experiment are collected in an *IQA database*. I divided IQA databases into two groups depending on the distortions applied to the reference images: conventional distortions (see Section 4.3) and gamut-

mapping distortions (see Section 4.4). The *Combined Gamut-Mapping Database* proposed in Section 4.5 is a collection of existing gamut-mapping datasets with reliable subjective data comprising 277 reference images, 2479 distorted images, and almost 30 000 subjective evaluations.

The prediction performance of IDMs is usually tested on IQA databases. Even though state-of-the-art IDMs show a high correlation to human judgments on conventional distortions, they are far away from perfect prediction on gamut-mapping distortions (see Table 5.3). An explanation of this discrepancy is that color is important for predicting gamut-mapping distortions but not for conventional distortions, and most IDMs only work on the monochromatic intensity level (see Section 5.2.1). In addition to missing color information, disregarding the viewing conditions may also lead to a decrease of the prediction performance (see Section 5.2.2).

In Section 5.3, a modular IDM framework was described which accounts for viewing conditions and color information. The viewing conditions are considered by a so-called *image-appearance model* (IAM) which is used to normalize the images to standard viewing conditions. The normalized images are represented in a working color space which provides access to color attributes, namely lightness, chroma, and hue. From these attributes, so-called *image-difference features* (IDFs) are extracted – characteristic values for local image differences with respect to human perception. In the last step of the framework, the IDFs are combined into an IDM denoted as *Color-Image-Difference* (CID) metric. The specific working color space, IAM, and parameter determination were elaborated in Section 5.4.

The promising prediction performance of the CID metric was verified in Section 5.5. On gamut-mapping distortions, CID shows the best prediction performance of the investigated IDMs – a significant improvement of $\approx 20\%$ on the hit-rate ratio (defined in Section 4.2.3) compared with the second best IDM. The incorporation of viewing conditions and color information seems to have a positive influence on the prediction accuracy. On conventional distortions, the impact of color is low but chromatic IDFs do not adversely affect CID's high correlation to human judgments – albeit not the highest correlation of the investigated IDMs.

An IDM which agrees with the population-average image-difference perception offers a variety of applications if used as an objective function. I.e., in an optimization problem we are seeking for an image which has a minimal image-difference prediction to the reference image and is thus perceived to be the best reproduction of the reference image. Gamut mapping can be formulated as such an optimization problem constrained to in-gamut images (see Equation 6.2). Since CID has a variable structure and showed the best prediction performance on gamut-mapping distortions, it was used as an objective function for optimizing gamut mapping denoted as *CID-based gamut-mapping optimization* (see Section 6.3). A visual experiment confirmed CID's promising performance as objective function for gamut mapping on a very small newspaper gamut – 83 % of the subjects preferred CID-based gamut-mapping optimization to two common gamut-mapping algorithms.

Even though the performance of CID-based gamut-mapping optimization was high, clearly visible artifacts occurred at the optimized images, in particular lightness inversion, chromatic ringing, chromatic edges, and lightness banding. The reason why these artifacts are created is that the prediction of the CID metric does not fully agree with human perception. Therefore, the CID metric was modified and its parameters were adjusted so as to yield artifact-free optimization results by addressing each of the artifacts (see Section 6.4). The modified metric is called *improved Color-Image-Difference* (iCID) metric which did not require parameter training on an IQA database – as opposed to most IDMs and in particular the CID metric.

This peculiarity of the iCID metric – adjustment of iCID by yielding artifact-free optimization results – might lead to an adverse effect on the prediction performance. Therefore, the performance of iCID was tested on gamut-mapping distortions (see Section 6.5.2) as well as on conventional distortions (see Section 6.5.3). While CID and iCID do not show a significant difference on gamut-mapping distortions, iCID outperforms CID on conventional distortions and nearly reaches the prediction accuracy of the best performing IDM which, however, has a much worse prediction performance on gamut-mapping distortions.

The by-product of the derivation of iCID – iCID-based gamut-mapping optimization – was investigated in Section 6.5.4. The optimized images benefit from the preservation of local contrast, structure, and color of the original image to a great extent. iCID-based gamut-mapping optimization was compared with a common as well as with a state-of-the-art gamut-mapping transformation in a visual experiment including 15 subjects. A significant majority preferred the iCID-based optimization on the used small newspaper gamut. Depending on the desired color attributes for the optimized images, iCID-based gamut-mapping optimization allows to choose between so-called *optimization intents* defined in Section 6.5.5.

In addition to gamut mapping, another application of iCID is presented in this work: *high-dynamic-range* (HDR) imaging (described in Section 7.2). The standard reproduction workflow of an HDR image is to first transform the image into a *low-dynamic-range* (LDR) representation by a tone-mapping operator and then to transform the LDR image into the output device's gamut by a gamut-mapping algorithm. To be able to perform the transformation of HDR images in a single step – called *HDR gamut mapping* – the concept of perceptual HDR color spaces was presented in Section 7.3. HDR gamut mapping can be performed like color (LDR) gamut mapping, but not many investigations of gamut mapping in the scope of HDR were made so far.

The proposed concept of HDR gamut mapping as well as the proposed perceptual HDR color space were investigated by employing iCID-based gamut-mapping optimization (see Section 7.4). This optimization – adapted to HDR – required a representation of the images in the perceptual HDR color space. The perceived image difference between the HDR image and a tone- and gamut-mapped image was minimized by using iCID as the objective function. The results of the HDR optimization were viewed on an HDR display (see Section 7.5). Contrast ratios and structural information of the HDR images were

preserved to a great extent indicating that the proposed concept of HDR gamut mapping is feasible and that the proposed HDR color space is suitable for HDR gamut mapping (see Section 7.6).

8.2 Outlook

This section presents – in brevity and without a relation between the single topics – suggestions for further investigations and applications of the presented work.

The iCID metric already shows the best prediction performance on gamut-mapping distortions and is close to the best prediction performance on conventional distortions. To further improve iCID's assessment of image differences, the following modifications may be beneficial for iCID:

- Applying a multi-scale approach to iCID like done for CID (only on conventional distortions).
- Improving image normalization, e.g., by applying a more suitable image-appearance model or, at least, more suitable contrast-sensitivity functions (for instance, orientation-dependent contrast-sensitivity functions).
- Changing the pooling strategy (up to now the mean), e.g., averaging by saliency maps.
- Including semantics into the metric, e.g., by face detection.

Actually, in a yet unpublished paper by Le Moan et al. [70], a multi-scale version of the iCID metric denoted as *multi-scale iCID* was proposed. It almost yields the highest correlation to human judgments on the largest available image-quality-assessment database. In particular, multi-scale iCID (Spearman correlation of 0.861) outperforms FSIMc (Spearman correlation of 0.851) on the TID2013 database where FSIMc was up to now the best performing image-difference metric (see Section 6.5.3).

The high performance of iCID-based gamut-mapping optimization was verified on an LCD display showing the optimized images as sRGB representations. However, the optimization was not tested on printed hardcopies. Problems may occur at the separation transformation, i.e., the transformation that maps the gamut-mapped image to the ink values of the printer – in particular, at colors near the darkest point of the color gamut (black point) or at smooth color gradients. Moreover, uniform areas that are bigger than iCID's sliding window side length k will remain uniform for most gamut-mapping transformations but need not be uniform for iCID-based gamut-mapping optimization. This problem was not evident on the display for the used images but may be disturbing on printed hardcopies. I suggest to post-process the optimized images to meet the requirements of the printing process. To conclude, a thorough investigation of iCID-based gamut-mapping optimization with respect to printing should be considered in future work.

Most image-difference metrics do not consider the viewing conditions. They assume an office environment – i.e., regarding the images on a standard display standing on the desk in an office – which is sufficient for many applications. In a cinema environment, however, the viewing conditions are quite different – such as the visual resolution or the dim illumination. The viewing conditions are also different for mobile devices, especially if used outside in bright sunlight. In such cases, the viewing conditions are supposed to be considered for image-difference prediction like included in the iCID metric.

An example of an application of image-difference metrics is *multispectral imaging*. The iCID metric is defined on colorimetrically specified images. It is based on hypotheses about human perception and takes into account the viewing conditions. A different type of images are *multispectral images* which have more than three color channels (each channel is defined by its spectral sensitivity). Thus, multispectral images contain more information and are less dependent on viewing conditions. To assess the image difference between the original spectral image and its reproduction, several spectral image-difference metrics were proposed. These spectral metrics, however, mostly do not take into account human perception [71]. Therefore, Le Moan and Urban proposed a spectral image-difference metric based on iCID denoted as the *Spectral Image Difference* (SID) metric [71]. SID renders the multispectral images for a variety of illuminants and predicts their image differences for each illuminant using iCID. Finally, the predictions for the different illuminants are pooled to a single spectral image-difference prediction. Further applications of image-difference metrics in recent image-processing tasks may include 3D-image quality assessment or gamut expansion.

The concept of high-dynamic-range gamut mapping opens up new prospects in the emerging field of high-dynamic-range imaging. Diminishing the standard workflow to a single transformation may retain color contrast ratios and allow for faster processing of high-dynamic-range images. In particular, a look-up table encoding of high-dynamic-range gamut mapping is supposed to allow real-time applications such as monitoring. Moreover, high-dynamic-range imaging could be included to common color-management workflows by introducing high-dynamic-range gamut mapping as an extension of ICC color profiles.

8.3 Scientific Added Value

The upcoming list designates the main scientific outcome emerged during my doctorate and presented in this thesis:

- Extension of the hit-rate concept by the hit-rate ratio and the significance analysis.
- A thorough investigation of state-of-the-art image-difference metrics on a large gamut-mapping database.

- Creation of the CID metric which yields the best prediction performance for gamut-mapping distortions.
- Observation that on gamut-mapping distortions, image-difference features extracted from different scales highly correlate.
- Improvement of the CID metric denoted as iCID which has an even higher prediction accuracy on conventional distortions.
- Parameter adjustment by improving the results if the image-difference metric is used as an objective function for optimization tasks – instead of training on an image-quality-assessment database.
- Introduction of iCID-based gamut-mapping optimization which significantly outperforms a state-of-the-art spatial gamut-mapping algorithm on a small newspaper gamut.
- Introduction of optimization intents by varying the parameters of the objective function within the optimization.
- Adaptation of an existing high-dynamic-range color space to an almost perceptually uniform and cross-contamination free color space called hdr-LAB2000HL.
- The concept of transforming high-dynamic-range images into low-dynamic-range devices' gamuts in a single step called high-dynamic-range gamut mapping.

References

[1] S. A. Ajagamelle, M. Pedersen, and G. Simone. Analysis of the difference of gaussians model in image difference metrics. In *CGIV 2010 – 5th European Conference on Colour in Graphics, Imaging, and Vision*, pp. 489–496, 2010.

[2] A. O. Akyüz and E. Reinhard. Color appearance in high-dynamic-range imaging. *Journal of Electronic Imaging*, 15(3):033001, 2006.

[3] A. Alsam and I. Farup. Colour Gamut Mapping as a Constrained Variational Problem. In *Image Analysis*, Vol. 5575, pp. 109–118. Springer, 2009.

[4] A. Alsam and I. Farup. Spatial colour gamut mapping by means of anisotropic diffusion. In *Computational Color Imaging*, pp. 113–124. Springer, 2011.

[5] İ. Avcıbaş, B. Sankur, and K. Sayood. Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, 11(2):206–223, 2002.

[6] R. Bala, R. deQueiroz, R. Eschbach, and W. Wu. Gamut Mapping to Preserve Spatial Luminance Variations. *Journal of Imaging Science and Technology*, 45(5):436–443, 2001.

[7] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice*. A K Peters, 2011.

[8] Z. Barańczuk, P. Zolliker, and J. Giesen. Image-Individualized Gamut Mapping Algorithms. *Journal of Imaging Science and Technology*, 54(3):030201, 2010.

[9] P. G. J. Barten. Formula for the contrast sensitivity of the human eye. In *Proc. SPIE 5294, Image Quality and System Performance*, pp. 231–238, 2004.

[10] R. S. Berns. *Billmeyer and Saltzman's Principles of Color Technology*. John Wiley & Sons, 3rd edition, 2000.

[11] böhringer friedrich. *Tropfen auf Gras*, Aug 2009. [Accessed: Apr-02-2013]. <http://commons.wikimedia.org/wiki/File:Tropfen_auf_Gras_2.JPG>

[12] H. R. Blackwell. Contrast Thresholds of the Human Eye. *Journal of the Optical Society of America*, 36(11):624–643, 1946.

[13] N. Bonnier, F. Schmitt, and H. Brettel. Evaluation of Spatial Gamut Mapping Algorithms. In *IS&T/SID, 14th Color and Imaging Conference*, pp. 56–61, 2006.

[14] D. O. Bowker. Spatial frequency discrimination thresholds in different orientations. *Journal of the Optical Society of America*, 70(4):462–463, 1980.

[15] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3–4):324–345, 1952.

[16] G. J. Braun and M. D. Fairchild. General-Purpose Gamut-Mapping Algorithms: Evaluation of Contrast-Preserving Rescaling Functions for Color Gamut Mapping. In *Color and Imaging Conference*, Vol. 1999, pp. 167–172, 1999.

[17] L. Brown and X. Li. Confidence intervals for two sample binomial distribution. *Journal of Statistical Planning and Inference*, 130(1–2):359–375, 2005.

[18] F. W. Campbell and J. G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197(3):551–566, 1968.

[19] D. M. Chandler. Seven Challenges in Image Quality Assessment: Past, Present, and Future Research. *ISRN Signal Processing*, 2013, 2013.

[20] D. M. Chandler and S. S. Hemami. VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. *IEEE Transactions on Image Processing*, 16(9):2284–2298, 2007.

[21] S. Chen, A Beghdadi, and A Chetouani. Color Image Assessment Using Spatial Extension to CIE DE2000. In *International Conference on Consumer Electronics (ICCE 2008)*, pp. 1–2, 2008.

[22] CIE International Commission on Illumination. Parametric effects in colour-difference evaluation. Technical report, CIE 101-1993, 1993.

[23] CIE International Commission on Illumination. Improvement to Industrial colour-Difference Evaluation. Technical report, CIE 142-2001, 2001.

[24] CIE International Commission on Illumination. A Colour Appearance Model for Colour Management Systems: CIECAM02. Technical report, CIE 159:2004, 2004.

[25] CIE International Commission on Illumination. Colorimetry, 3rd Edition. Technical report, CIE 015:2004, 2004.

[26] G. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt. Uniform colour spaces based on the DIN99 colour-difference formula. *Color Research & Application*, 27(4):282–290, 2002.

[27] S. Daly. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity. In *Digital Images and Human Vision*, pp. 179–206. MIT Press, 1993.

[28] S. J. Daly. Application of a noise-adaptive contrast sensitivity function to image data compression. *Optical Engineering*, 29(8):977–987, 1990.

[29] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik. Image Quality Assessment Based on a Degradation Model. *IEEE Transactions on Image Processing*, 9(4):636–650, 2000.

[30] R. L. De Valois and K. K. De Valois. *Spatial Vision*. Oxford University Press, 1990.

[31] F. Dugay, I. Farup, and J. Y. Hardeberg. Perceptual Evaluation of Color Gamut Mapping Algorithms. *Color Research & Application*, 33(6):470–476, 2008.

[32] F. Ebner and M. D. Fairchild. Development and Testing of a Color Space (IPT) with Improved Hue Uniformity. In *IS&T/SID, 6th Color and Imaging Conference*, pp. 8–13, 1998.

[33] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. New full-reference quality metrics based on HVS. In *Second International Workshop on Video Processing and Quality Metrics*, 2006.

[34] EMPA Media Technology. *Gamut mapping data web page*. [Accessed: Feb-19-2014]. <http://empamedia.ethz.ch/gmdb/index.php>

[35] P. G. Engeldrum. Image Quality Modeling: Where Are We? In *IS&T, PICS 1999: Image Processing, Image Quality, Image Capture, Systems Conference*, pp. 251–255, 1999.

[36] P. G. Engeldrum. *Psychometric Scaling: A Toolkit For Imaging Systems Development*. Imcotek Press, 1st edition, 2000.

[37] U. Engelke, M. Kusuma, H.-J. Zepernick, and M. Caldera. Reduced-reference metric design for objective perceptual quality assessment in wireless imaging. *Signal Processing: Image Communication*, 24(7):525–547, 2009.

[38] M. D. Fairchild. *Color Appearance Models*. John Wiley & Sons, 2005.

[39] M. D. Fairchild. The HDR Photographic Survey. In *IS&T/SID, 15th Color and Imaging Conference*, pp. 233–238, 2007.

[40] M. D. Fairchild and P.-H. Chen. Brightness, lightness, and specifying color in high-dynamic-range scenes and images. In *Proc. SPIE 7867, Image Quality and System Performance VIII*, p. 78670O, 2011.

[41] M. D. Fairchild and D. R. Wyble. hdr-CIELAB and hdr-IPT: Simple Models for Describing the Color of High-Dynamic-Range and Wide-Color-Gamut Images. In *IS&T/SID, 18th Color and Imaging Conference*, pp. 322–326, 2010.

[42] I. Farup, C. Gatta, and A. Rizzi. A multiscale framework for spatial gamut mapping. *IEEE Transactions on Image Processing*, 16(10):2423–2435, 2007.

[43] G. T. Fechner. *Elemente der Psychophysik*, Vol. 1–2. Leipzig: Breitkopf und Härtel, 1860.

[44] J. A. Ferwerda and S. Luka. A high resolution high dynamic range display for vision research. *Journal of Vision*, 9(8):346a, 2009.

[45] R. Franzen. *Kodak Lossless True Color Image Suite.* [Accessed: Jan-18-2013]. <http://r0k.us/graphics/kodak/>

[46] M. Gaubatz. *MeTriX MuX Visual Quality Assessment Package.* [Accessed: Apr-15-2011]. <http://foulard.ece.cornell.edu/gaubatz/metrix_mux/>

[47] G. A. Gescheider. *Psychophysics: the fundamentals.* Lawrence Erlbaum Associates, 3rd edition, 1997.

[48] P. Getreuer. *Colorspace Package.* [Accessed: Mar-20-2014]. <http://www.getreuer.info/home/colorspace>

[49] J. Giesen, E. Schuberth, K. Simon, P. Zolliker, and O. Zweifel. Image-Dependent Gamut Mapping as Optimization Problem. *IEEE Transactions on Image Processing*, 16(10):2401–2410, 2007.

[50] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing Using MATLAB.* Prentice Hall, 2004.

[51] N. V. S. Graham. *Visual Pattern Analyzers.* Oxford University Press, 1989.

[52] J. Y. Hardeberg, E. Bando, and M. Pedersen. Evaluating colour image difference metrics for gamut-mapped images. *Coloration Technology*, 124(4):243–253, 2008.

[53] P.-C. Hung and R. S. Berns. Determination of constant Hue Loci for a CRT gamut and their predictions using color appearance spaces. *Color Research & Application*, 20(5):285–295, 1995.

[54] R. W. G. Hunt. *The Reproduction of Colour.* John Wiley & Sons, 6th edition, 2004.

[55] International Color Consortium (ICC). *Specification of Color Profiles*, 4.3.0.0 edition. Manual, 2010. <http://www.color.org>

[56] International Imaging Industry Association (I3A). CPIQ Initiative Phase 1 White Paper: Fundamentals and review of considered test methods. Technical report, 2007.

[57] R. E. Jacobson. An evaluation of image quality metrics. *Journal of Photographic Science*, 43(1):7–16, 1995.

[58] G. M. Johnson. *Measuring Images: Differences, Quality and Appearance.* PhD thesis, Rochester Institute of Technology, USA, 2003.

[59] G. M. Johnson and M. D. Fairchild. Darwinism of Color Image Difference Models. In *IS&T/SID, 9th Color and Imaging Conference*, pp. 108–112, 2001.

[60] jonathangarcia. *Color Balance*, Jul 2009. [Accessed: Mar-28-2013]. <http://www.fotopedia.com/items/flickr-3748383024>

[61] D. B. Judd. Ideal color space: Curvature of color space and its implications for industrial color tolerances. *Palette*, 29:25–31, 1968.

[62] B. W. Keelan. *Handbook of Image Quality: Characterization and Prediction.* Marcel Dekker, 2002.

[63] M. G. Kendall. A New Measure of Rank Correlation. *Biometrika*, 30(1–2):81–89, 1938.

[64] K. J. Kim, R. Mantiuk, and K. H. Lee. Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation luminance. In *Proc. SPIE 8651, Human Vision and Electronic Imaging XVIII*, p. 86511A, 2013.

[65] R. Kimmel, D. Shaked, M. Elad, and I. Sobel. Space-Dependent Color Gamut Mapping: A Variational Approach. *IEEE Transactions on Image Processing*, 14(6):796–803, 2005.

[66] J. Kuang, G. M. Johnson, and M. D. Fairchild. iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414, 2007.

[67] R. G. Kuehni. *Color Space and Its Divisions: Color Order from Antiquity to the Present.* John Wiley & Sons, 2003.

[68] E. C. Larson and D. M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 2010.

[69] P. Le Callet and F. Autrusseau. *Subjective quality assessment IRCCyN/IVC database*, 2005. [Accessed: Jan-17-2014]. <http://www.irccyn.ec-nantes.fr/ivcdb/>

[70] S. Le Moan, J. Preiss, and P. Urban. Evaluating the multi-scale iCID metric. In *Electronic Imaging 2015: Image Quality and System Performance XII (accepted)*, 2015.

[71] S. Le Moan and P. Urban. Image-Difference Prediction: From Color to Spectral. *IEEE Transactions on Image Processing*, 23(5):2058–2068, 2014.

[72] G. E. Legge and J. M. Foley. Contrast masking in human vision. *Journal of the Optical Society of America*, 70(12):1458–1471, 1980.

[73] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi. The Similarity Metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, 2004.

[74] I. Lissner. *Improving Color-Difference Formulas and Perceptual Color Spaces*. PhD thesis, Technische Universität Darmstadt, Germany, 2013.

[75] I. Lissner, J. Preiss, and P. Urban. Predicting Image Differences Based on Image-Difference Features. In *IS&T/SID, 19th Color and Imaging Conference*, pp. 23–28, 2011.

[76] I. Lissner, J. Preiss, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. Image-Difference Prediction: From Grayscale to Color. *IEEE Transactions on Image Processing*, 22(2):435–446, 2013.

[77] I. Lissner and P. Urban. How Perceptually Uniform Can a Hue Linear Color Space Be? In *IS&T/SID, 18th Color and Imaging Conference*, pp. 97–102, 2010.

[78] I. Lissner and P. Urban. Toward a Unified Color Space for Perception-Based Image Processing. *IEEE Transactions on Image Processing*, 21:1153–1168, 2012.

[79] M. R. Luo, G. Cui, and B. Rigg. The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000. *Color Research and Application*, 26(5):340–350, 2001.

[80] J. Mannos and D. Sakrison. The Effects of a Visual Fidelity Criterion of the Encoding of Images. *IEEE Transactions on Information Theory*, 20(4):525–536, 1974.

[81] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel. High Dynamic Range Image and Video Compression - Fidelity Matching Human Visual Performance. In *IEEE International Conference on Image Processing, 2007. ICIP 2007.*, Vol. 1, pp. I–9–I–12, 2007.

[82] R. Mantiuk, A Tomaszewska, and W Heidrich. Color correction for tone mapping. In *Computer Graphics Forum*, Vol. 28, pp. 193–202, 2009.

[83] mattbuck. *Bristol Balloon Fiesta*, Aug 2009. [Accessed: Apr-02-2013]. <http://commons.wikimedia.org/wiki/File:Bristol_Balloon_Fiesta_2009_MMB_04_G-OFXP.jpg>

[84] J. J. McCann. Color gamut mapping using spatial comparisons. In *Proc. SPIE 4300, Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts VI*, p. 126, 2001.

[85] T. Mitsa and K. L. Varkur. Evaluation of Contrast Sensitivity Functions for the Formulation of Quality Measures Incorporated in Halftoning Algorithms. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 301–304, 1993.

[86] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman. The CIECAM02 Color Appearance Model. In *IS&T/SID, 10th Color and Imaging Conference*, Vol. 2002, pp. 23–27, 2002.

[87] J. Morovič. *To Develop a Universal Gamut Mapping Algorithm*. PhD thesis, University of Derby, UK, 1998.

[88] J. Morovič. *Color Gamut Mapping*. John Wiley & Sons, 2008.

[89] J. Morovič and Y. Wang. A multi-resolution full-colour spatial gamut mapping algorithm. In *IS&T/SID, 11th Color and Imaging Conference*, pp. 282–287, 2003.

[90] J. H. Morrissey. New Method for the Assignment of Psychometric Scale Values from Incomplete Paired Comparison. *Journal of the Optical Society of America*, 45(5):373–378, 1955.

[91] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *Journal of Physiology*, 359:381–400, 1985.

[92] K. I. Naka and W. A. H. Rushton. S-potentials from colour units in the retina of fish (Cyprinidae). *Journal of Physiology*, 185(3):536–555, 1966.

[93] S. Nakauchi, S. Hatanaka, and S. Usui. Color Gamut Mapping Based on a Perceptual Image Difference Measure. *Color Research and Application*, 24(4):280–291, 1999.

[94] M. Narwaria, M. Perreira Da Silva, P. Le Callet, and R. Pepion. On improving the pooling in HDR-VDP-2 towards better HDR perceptual quality assessment. In *Proc. SPIE 9014, Human Vision and Electronic Imaging XIX*, p. 90140N, 2014.

[95] C. Oleari, M. Melgosa, and R. Huertas. Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space. *Journal of the Optical Society of America A*, 26(1):121–134, 2009.

[96] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita. *Image Quality Evaluation Database.* [Accessed: Jan-17-2014]. <http://mict.eng.u-toyama.ac.jp/mictdb.html>

[97] S. N. Pattanaik, M. D. Fairchild, J. A. Ferwerda, and D. P. Greenberg. Multiscale Model of Adaptation, Spatial Vision and Color Appearance. In *IS&T/SID, 6th Color and Imaging Conference*, pp. 2–7, 1998.

[98] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg. A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 287–298, 1998.

[99] M. Pedersen. *Image quality metrics for the evaluation of printing workflows.* PhD thesis, University of Oslo, Norway, 2011.

[100] M. Pedersen and J. Y. Hardeberg. Survey of full-reference image quality metrics. *Høgskolen i Gjøviks rapportserie*, (5), 2009.

[101] M. Pedersen and J. Y. Hardeberg. Full-Reference Image Quality Metrics: Classification and Evaluation. *Foundations and Trends in Computer Graphics and Vision*, 7(1):1–80, 2012. <http://dx.doi.org/10.1561/0600000037>

[102] E. Peli, L. E. Arend, G. M. Young, and R. B. Goldstein. Contrast sensitivity to patch stimuli: Effects of spatial bandwidth and temporal presentation. *Spatial Vision*, 7(1):1–14, 1993.

[103] A. B. Poirson and B. A. Wandell. Appearance of colored patterns: pattern-color separability. *Journal of the Optical Society of America A*, 10(12):2458–2470, 1993.

[104] A. B. Poirson and B. A. Wandell. Pattern-Color Separable Pathways Predict Sensitivity to Simple Colored Patterns. *Vision Research*, 36(4):515–526, 1996.

[105] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin. Metrics Performance Comparison for Color Image Database. In *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, p. 6, Scottsdale, Ariz., Jan. 14–16 2009.

[106] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli. Modified Image Visual Quality Metrics for Contrast Change and Mean Shift Accounting. In *11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 2011*, pp. 305–311, 2011.

[107] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo. Color Image Database TID2013: Peculiarities and Preliminary Results. In *4th European Workshop on Visual Information Processing EUVIP2013, June 10-12, Paris, France*, pp. 106–111, 2013.

[108] N. Ponomarenko, O. Ieremeiev, V. Lukin, L. Jin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo. A New Color Image Database TID2013: Innovations and Results. In *Advanced Concepts for Intelligent Vision Systems*, Vol. 8192 of *Lecture Notes in Computer Science*, pp. 402–413. Springer, 2013.

[109] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics. *Advances of Modern Radioelectronics*, 10:30–45, 2009.

[110] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin. On between-coefficient contrast masking of DCT basis functions. In *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07*, 2007.

[111] J. Preiss, M. D. Fairchild, J. A. Ferwerda, and P. Urban. Gamut Mapping in a High-Dynamic-Range Color Space. In *Proc. SPIE 9015, Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications*, p. 90150A, 2014.

[112] J. Preiss, F. Fernandes, and P. Urban. Color-Image Quality Assessment: From Prediction to Optimization. *IEEE Transactions on Image Processing*, 23(3):1366–1378, 2014.

[113] J. Preiss, I. Lissner, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. The Impact of Image-Difference Features on Perceived Image Differences. In *CGIV 2012 – 6th European Conference on Colour in Graphics, Imaging, and Vision*, pp. 43–48, 2012.

[114] J. Preiss and P. Urban. Information Content Redundancies of Image Quality Assessments. In *17. Workshop Farbbildverarbeitung*, pp. 97–105, 2011.

[115] J. Preiss and P. Urban. Image-Difference Measure Optimized Gamut Mapping. In *IS&T/SID, 20th Color and Imaging Conference*, pp. 230–235, 2012.

[116] E. Reinhard. Tone Reproduction and Color Appearance Modeling: Two Sides of the Same Coin? In *IS&T/SID, 19th Color and Imaging Conference*, pp. 171–176, 2011.

[117] E. Reinhard, E. A. Khan, A. O. Akyüz, and G. M. Johnson. *Color Imaging: Fundamentals and Applications*. A K Peters, 2008.

[118] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging*. Morgan Kaufmann Publishers, 2006.

[119] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao. Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A*, 15(8):2036–2045, 1998.

[120] O. H. Schade, Sr. Optical and Photoelectric Analog of the Eye. *Journal of the Optical Society of America*, 46(9):721–738, 1956.

[121] M. Scheller Lichtenauer, P. Zolliker, I. Lissner, J. Preiss, and P. Urban. Learning Image Similarity Measures from Choice Data. In *CGIV 2012 – 6th European Conference on Colour in Graphics, Imaging, and Vision*, pp. 24–30, 2012.

[122] M. Scheller Lichtenauer, P. Zolliker, J. Preiss, and P. Urban. Image similarity for chromatic content. In *18. Workshop Farbbildverarbeitung*, pp. 107–117, 2012.

[123] C. Schlick. Quantization Techniques for Visualization of High Dynamic Range Pictures. In *Photorealistic Rendering Techniques*, pp. 7–20. Springer, 1995.

[124] S. Schlotzhauer. *Elementary Statistics Using JMP*. SAS Institute, 2007.

[125] K. Seshadrinathan and A. C. Bovik. Advances in Image and Video Quality Assessment. In *Encyclopedia of Multimedia*, pp. 8–17. Springer, 2nd edition, 2008.

[126] G. Sharma. *Digital Color Imaging Handbook*. CRC Press, 2002.

[127] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[128] H. R. Sheikh, A. C. Bovik, and G. de Veciana. An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005.

[129] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, 15(11):3441–3452, 2006.

[130] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database Release 2*. [Accessed: Jan-15-2014]. <http://live.ece.utexas.edu/research/quality/>

[131] M. Shohara and K. Kotani. The dependence of visual noise perception on background color and luminance. In *Picture Coding Symposium (PCS)*, pp. 594–597, 2010.

[132] G. Simone, C. Oleari, and I. Farup. Performance of the euclidean color-difference formula in log-compressed OSA-UCS space applied to modified-image-difference metrics. In *11th Congress of the International Colour Association (AIC)*, p. 81, 2009.

[133] Simple Insomnia. *Berber camel driver*, Sep 2009. [Accessed: Apr-02-2013]. <http://commons.wikimedia.org/wiki/File:S1000RR_Composing.jpg>

[134] X. Song, G. M. Johnson, and M. D. Fairchild. Minimizing the Perception of Chromatic Noise in Digital Images. In *IS&T/SID, 12th Color and Imaging Conference*, pp. 340–346, 2004.

[135] C. Spearman. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15:72–101, 1904.

[136] Stefan Krause. *S1000RR*, Oct 2010. [Accessed: Apr-02-2013]. <http://commons.wikimedia.org/wiki/File:S1000RR_Composing.jpg>

[137] S. S. Stevens. To Honor Fechner and Repeal His Law. *Science*, 133(3446):80–86, 1961.

[138] Y. Tadmor and D. J. Tolhurst. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Research*, 40(22):3145–3157, 2000.

[139] L. A. Taplin and G. M. Johnson. When good hues go bad. In *CGIV2004 – Second European Conference on Colour in Graphics, Imaging, and Vision*, pp. 348–352, 2004.

[140] K.-H. Thung and P. Raveendran. A survey of image quality measures. In *International Conference for Technical Postgraduates (TECHPOS)*, pp. 1–4, 2009.

[141] L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927.

[142] A. Toet and M. P. Lucassen. A new universal colour image fidelity metric. *Displays*, 24:197–207, 2003.

[143] P. Urban, D. Schleicher, M. R. Rosen, and R. S. Berns. Embedding non-Euclidean color spaces into Euclidean color spaces with minimal isometric disagreement. *Journal of the Optical Society of America A*, 24(6):1516–1528, 2007.

[144] B. A. Wandell. *Foundations of Vision*. Sinauer Associates, 1995.

[145] Z. Wang. *SSIM*, 2009. [Accessed: May-13-2013]. <https://ece.uwaterloo.ca/~z70wang/research/ssim/ssim.m>

[146] Z. Wang and A. C. Bovik. A Universal Image Quality Index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.

[147] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[148] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009

[149] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600 – 612, April 2004.

[150] Z. Wang and Q. Li. Information Content Weighting for Perceptual Image Quality Assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, 2011.

[151] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2004*, Vol. 2, pp. 1398–1402, 2003.

[152] C. E. Wright and N. Drasdo. The influence of age on the spatial and temporal contrast sensitivity function. *Documenta Ophthalmologica*, 59(4):385–395, 1985.

[153] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, 2nd edition, 2000.

[154] A. Zaric, N. Tatalovic, N. Brajkovic, H. Hlevnjak, M. Loncaric, E. Dumic, and S. Grgic. VCL@FER Image Quality Assessment Database. In *ELMAR, 2011 Proceedings*, pp. 105–110, 2011.

[155] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, August 2011.

[156] X. Zhang and B. A. Wandell. A Spatial Extension of CIELAB for Digital Color Image Reproduction. *Society for Information Display*, 5:61–63, March 1997.

[157] P. Zolliker, Z. Barańczuk, and J. Giesen. Image Fusion for optimizing Gamut Mapping. In *IS&T/SID, 19th Color and Imaging Conference*, pp. 109–114, 2011.

[158] P. Zolliker and K. Simon. Retaining Local Image Information in Gamut Mapping Algorithms. *IEEE Transactions on Image Processing*, 16(3):664–672, 2007.

[159] Z. Zolliker, P. Barańczuk, I. Sprow, and J. Giesen. Conjoint Analysis for Evaluating Parameterized Gamut Mapping Algorithms. *IEEE Transactions on Image Processing*, 19(3):758–769, 2010.

Appendix A

Mathematical Formulations

A.1 Specification of sRGB

A commonly used RGB color space (see Section 2.3.1) for images is sRGB, a *standardized RGB* (red-green-blue) color space encoded with 8 bit per color channel and non-linear gamma compression. All input images in this thesis are defined as sRGB images but they are further transformed into different working color spaces. The specification of sRGB is given in standard *IEC 61966-2-1:1999* which uses the primaries and the white point of standard *ITU-R BT.709-3*. Here, sRGB values are represented as normalized values of the three primaries $(R_{\mathrm{sRGB}}, G_{\mathrm{sRGB}}, B_{\mathrm{sRGB}})$ in the range [0,1].

Please note that I use MATLAB for image processing and sometimes use its built-in functions. If this is the case, the name of the used function is given with an explanation. For reading an sRGB image, I apply the *imread* function on the image's file format and then normalize the image's range to the interval [0,1]. Furthermore, I do not refer to a specific white point in color-space transformations, i.e., I do not consider illumination or standard observer. This is without loss of generality because changing the illumination can be done by chromatic adaptation, e.g., included in the *CIECAM02* color appearance model [86]. For the sake of completeness, most of the following transformations are based on illuminant CIE D65, CIE 2° standard observer.

A.2 Color-Space Transformations

A.2.1 sRGB to Grayscale Transformation

sRGB images consist of three color channels. In an opponent color space a color signal is separated into one achromatic channel and two chromatic channels. The achromatic channel refers to the intensity of the image which are grayscale values. Grayscale values of an image can be obtained by computing the luminance Y in a linear RGB color space or the luma Y' in a gamma-compressed RGB color space.

For the transformation from sRGB to grayscale values, I apply the *rgb2gray* built-in MATLAB function which uses the luminance transformation from RGB to NTSC (rounded):

$$Y' = 0.299 \cdot R_{\mathrm{sRGB}} + 0.587 \cdot G_{\mathrm{sRGB}} + 0.114 \cdot B_{\mathrm{sRGB}}. \tag{A.1}$$

The rounded transformation equals the luma Y' of standard *ITU-R BT.601-7*. Additionally, since sRGB is a gamma-compressed color space, I use the notation luma Y' for grayscale values.

A.2.2 sRGB to CIEXYZ Transformation

First, the gamma-compressed sRGB values are transformed into linear RGB values $(R_{\mathrm{RGB}}, G_{\mathrm{RGB}}, B_{\mathrm{RGB}})$ by a function similar to a power function with exponent 2.2. Then, a linear transformation to CIEXYZ values (see Section 2.3.2) is performed by simple matrix multiplication. The utilized transformations are

taken from the *Colorspace Package* by Getreuer [48] (rounded) which are the same as the transformations recommended in standard *IEC 61966-2-1:1999* except rounding errors:

$$(R_{\mathrm{RGB}}, G_{\mathrm{RGB}}, B_{\mathrm{RGB}}) = (f(R_{\mathrm{sRGB}}), f(G_{\mathrm{sRGB}}), f(B_{\mathrm{sRGB}})) \tag{A.2}$$

where function $f(\xi)$ is defined as

$$f(\xi) = \begin{cases} \xi/12.92 & \text{if } \xi \leq 0.040448 \\ [(\xi + 0.055)/1.055]^{2.4} & \text{else} \end{cases} . \tag{A.3}$$

and

$$\mathrm{X} = 0.4124 \cdot R_{\mathrm{RGB}} + 0.3576 \cdot G_{\mathrm{RGB}} + 0.1805 \cdot B_{\mathrm{RGB}} \tag{A.4}$$

$$\mathrm{Y} = 0.2126 \cdot R_{\mathrm{RGB}} + 0.7152 \cdot G_{\mathrm{RGB}} + 0.0722 \cdot B_{\mathrm{RGB}} \tag{A.5}$$

$$\mathrm{Z} = 0.0193 \cdot R_{\mathrm{RGB}} + 0.1192 \cdot G_{\mathrm{RGB}} + 0.9505 \cdot B_{\mathrm{RGB}}. \tag{A.6}$$

The reverse transformations from CIEXYZ to sRGB values are

$$R_{\mathrm{RGB}} = 3.2406 \cdot \mathrm{X} - 1.5372 \cdot \mathrm{Y} - 0.4986 \cdot \mathrm{Z} \tag{A.7}$$

$$G_{\mathrm{RGB}} = -0.9689 \cdot \mathrm{X} + 1.8758 \cdot \mathrm{Y} + 0.0415 \cdot \mathrm{Z} \tag{A.8}$$

$$B_{\mathrm{RGB}} = 0.0557 \cdot \mathrm{X} - 0.2040 \cdot \mathrm{Y} + 1.0570 \cdot \mathrm{Z} \tag{A.9}$$

and

$$(R_{\mathrm{sRGB}}, G_{\mathrm{sRGB}}, B_{\mathrm{sRGB}}) = (g(R_{\mathrm{RGB}}), g(G_{\mathrm{RGB}}), g(B_{\mathrm{RGB}})) \tag{A.10}$$

where function $g(\xi)$ is defined as

$$g(\xi) = \begin{cases} 12.92 \cdot \xi & \text{if } \xi \leq 0.00313067 \\ 1.055 \cdot \xi^{1/2.4} - 0.055 & \text{else} \end{cases} . \tag{A.11}$$

A.2.3 CIEXYZ to CIELAB Transformation

The transformations of the CIEXYZ tristimulus values $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ into the CIELAB color-opponent coordinates $(L^*, a^*, b^*$, see Section 2.3.3) with reference white point $(\mathrm{X}_n, \mathrm{Y}_n, \mathrm{Z}_n)$ are given in [25]. They coincide except rounding errors with the calculations within this thesis using the *Colorspace Package* by Getreuer [48]:

$$L^* = 116 f(\mathrm{Y}/\mathrm{Y}_n) - 16 \tag{A.12}$$

$$a^* = 500\,[f(\mathrm{X}/\mathrm{X}_n) - f(\mathrm{Y}/\mathrm{Y}_n)] \tag{A.13}$$

$$b^* = 200\,[f(\mathrm{Y}/\mathrm{Y}_n) - f(\mathrm{Z}/\mathrm{Z}_n)] \tag{A.14}$$

where function $f(\xi)$ is defined as

$$f(\xi) = \begin{cases} (841/108) \cdot \xi + 4/29 & \text{if } \xi \leq 0.008856 \\ \xi^{1/3} & \text{else} \end{cases} . \tag{A.15}$$

The revers transformation from CIELAB to CIEXYZ color space using the *Colorspace Package* with reference white point $(\mathrm{X}_n, \mathrm{Y}_n, \mathrm{Z}_n)$ goes:

$$\mathrm{X} = \mathrm{X}_n \cdot g((L^* + 16)/116 + a^*/500) \tag{A.16}$$

$$\mathrm{Y} = \mathrm{Y}_n \cdot g((L^* + 16)/116) \tag{A.17}$$

$$\mathrm{Z} = \mathrm{X}_n \cdot g((L^* + 16)/116 - b^*/200) \tag{A.18}$$

where function $g(\xi)$ is defined as

$$g(\xi) = \begin{cases} (108/841) \cdot (\xi - 4/29) & \text{if } \xi^3 \leq 0.008856 \\ \xi^3 & \text{else} \end{cases} . \tag{A.19}$$

The polar representation of a^* and b^* corresponds to the chroma predictor C^* and the hue predictor H^*. The transformation of (a^*, b^*) to (C^*, H^*) is given by [25]:

$$C^* = \sqrt{a^{*2} + b^{*2}} \tag{A.20}$$

$$H^* = \arctan(b^*/a^*) \tag{A.21}$$

where arctan is the *four-quadrant inverse tangent* called by the *atan2* MATLAB function.

The reverse transformation is

$$a^* = C^* \cdot \cos(H^*) \tag{A.22}$$

$$b^* = C^* \cdot \sin(H^*). \tag{A.23}$$

A.2.4 CIELAB to LAB2000HL Transformation

The transformation from CIELAB values (L^*, a^*, b^*) to LAB2000HL values $(L, a, b$, see Section 2.3.4) as well as the reverse transformation are performed by applying look-up tables given in the supplementary material of the LAB2000HL paper [78]. Look-up tables are arrays of pre-calculated values which are assigned to any input value directly or by inter- or extrapolation of the pre-calculated values.

A.2.5 CIEXYZ to hdr-LAB2000HL

The transformation into the hdr-LAB2000HL color space (see Section 7.3) starts with the transformation from CIEXYZ tristimulus values $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ with reference white point $(\mathrm{X}_n, \mathrm{Y}_n, \mathrm{Z}_n)$ into high-dynamic-range opponent-color coordinates $(L_{HDR}, a_{HDR}, b_{HDR})$ of the hdr-CIELAB high-dynamic-range color space [40]:

$$L_{HDR} = f(\mathrm{Y}/\mathrm{Y}_n) \tag{A.24}$$

$$a_{HDR} = 5[f(\mathrm{X}/\mathrm{X}_n) - f(\mathrm{Y}/\mathrm{Y}_n)] \tag{A.25}$$

$$b_{HDR} = 2[f(\mathrm{Y}/\mathrm{Y}_n) - f(\mathrm{Z}/\mathrm{Z}_n)] \tag{A.26}$$

where function $f(\xi)$ is defined as

$$f(\xi) = 247\frac{\xi^\epsilon}{\xi^\epsilon + 2^\epsilon} + 0.02 \tag{A.27}$$

with $\epsilon = 0.58$.

Then, L_{HDR}, a_{HDR}, and b_{HDR} are transformed into the hdr-LAB2000HL color space analogously to the CIELAB to LAB2000HL transformation (see Section A.2.4). For lightness values L_{HDR} above diffuse white ($L_{HDR} > 100$), the option *extrap* of the *interp1* MATLAB function is used.

The same holds for the reverse transformation from the hdr-LAB2000HL color space into the hdr-CIELAB color space. The reverse transformation from high-dynamic-range opponent-color coordinates into CIEXYZ tristimulus values with reference white point $(\mathrm{X}_n, \mathrm{Y}_n, \mathrm{Z}_n)$ is

$$\mathrm{X} = \mathrm{X}_n \cdot g(L_{HDR} + a_{HDR}/5) \tag{A.28}$$

$$\mathrm{Y} = \mathrm{Y}_n \cdot g(L_{HDR}) \tag{A.29}$$

$$\mathrm{Z} = \mathrm{Z}_n \cdot g(L_{HDR} - b_{HDR}/2) \tag{A.30}$$

where function $g(\xi)$ is defined as

$$g(\xi) = \left(\frac{(\xi - 0.02)\cdot 2^\epsilon}{247.02 - \xi}\right)^{1/\epsilon} \tag{A.31}$$

with $\epsilon = 0.58$.

A.3 SSIM Luminance in a Perceptually Uniform Color Space

The SSIM index (see Section 3.3.2) is calculated on grayscale images which represent the intensity of an image, i.e., the luminance Y in a linear RGB color space or the luma Y' in a gamma-compressed RGB color space (see Section A.2.1). Thus, the image-difference comparisons l, c, and s are applied in an intensity linear color space. The behavior of the SSIM luminance comparison $l(x, y)$ (see Equation 3.10) in such a color space is illustrated in Figure A.1. However, this is not valid in a perceptually uniform color space so that the luminance comparison l has to be reformulated [76]:

To emphasize the reference to intensity I denote in the following the local statistics mean μ_x and μ_y (see Equation 3.7) as Y_x and Y_y, respectively. The constant c_1 is neglected which affects the equation only for a denominator close to zero. The luminance comparison l (see Equation 3.10) then reads:

$$l(x, y) = \frac{2Y_x Y_y}{{Y_x}^2 + {Y_y}^2}. \tag{A.32}$$

According to *Fechner's law* [43] the relationship between intensity Y and perceptually uniform lightness L is logarithmic. The more correct Steven's law [137] suggests a power function but the difference to

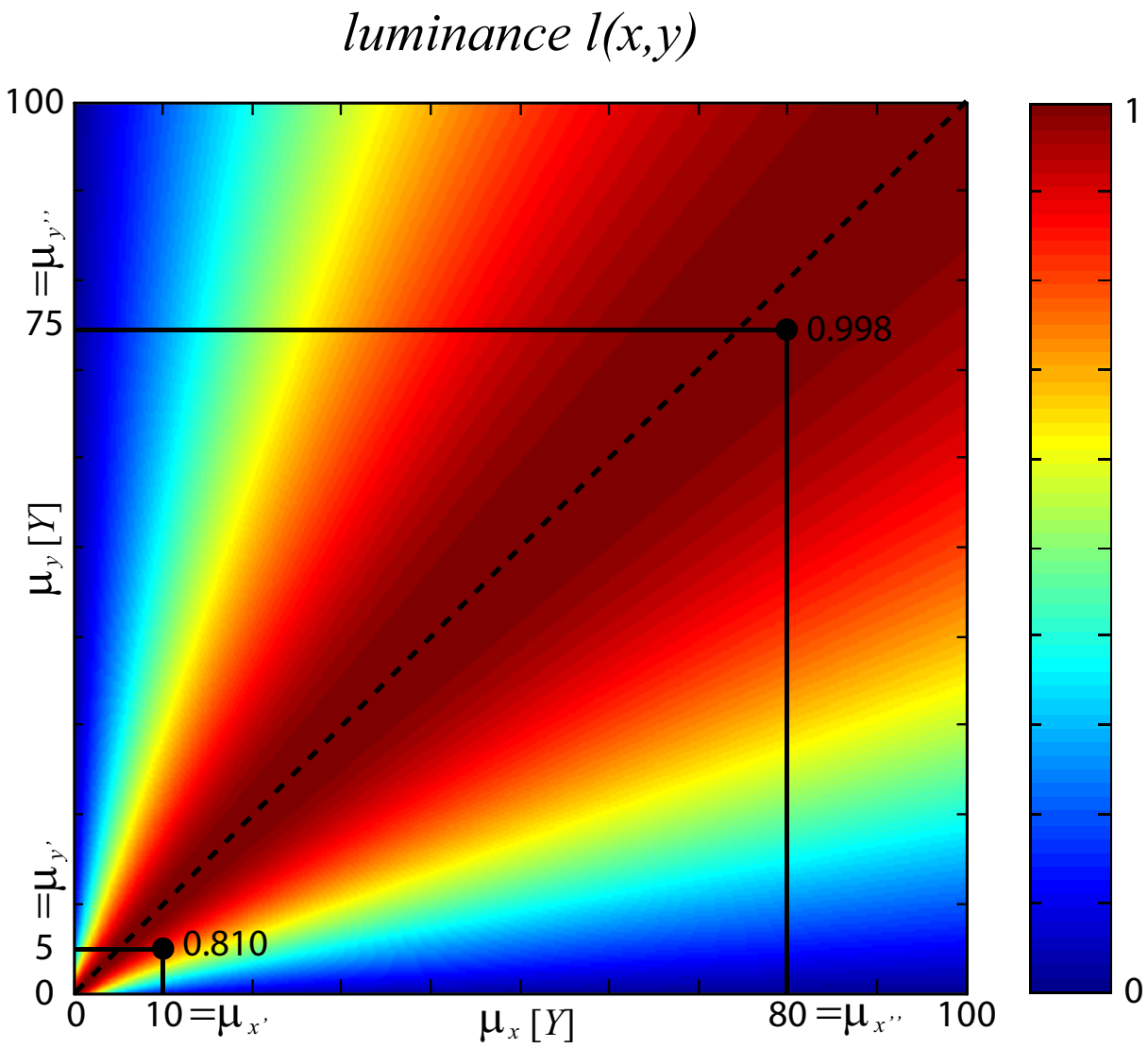


Figure A.1: Surface plot of the SSIM luminance comparison $l(x, y)$ with $c_1 = 6.5025$ projected onto the μ_x-μ_y-plane. As the function is used in an intensity linear color space μ_x and μ_y correspond to intensity Y. The luminance comparison increases with increasing intensity Y for constant intensity differences, e.g., $\Delta Y(\mu_{x'} = 10, \mu_{y'} = 5) = \Delta Y(\mu_{x''} = 80, \mu_{y''} = 75) = 5$ but $l(x', y') = 0.810$ and $l(x'', x'') = 0.998$. I.e., intensity differences are more disturbing in low intensity regions because smaller values of $l(x, y)$ correspond to larger perceived differences. A similar version of the figure has already been published in [113].

Fechner is rather small. Hence, Fechner's law is used here for the sake of simplicity:

$$L = L_0 \cdot \log \frac{Y}{Y_0} \tag{A.33}$$

with Y_0 representing the threshold of the just-not-noticeable perception of intensity Y and L_0 representing the increase of lightness L. Please note that my denotation corrects the wrong denotation from the corresponding paper [76] because Y_0 and L_0 do not refer to maximal values.

Inverting Equation A.33 for intensities Y_x and Y_y leads to:

$$Y_x = Y_0 \cdot \mathrm{e}^{L_x/L_0} \tag{A.34}$$

$$Y_y = Y_0 \cdot \mathrm{e}^{L_y/L_0} \tag{A.35}$$

which is substituted into Equation A.32:

$$l_L(x,y) = \frac{2\mathrm{e}^{L_x/L_0}\mathrm{e}^{L_y/L_0}}{\mathrm{e}^{2L_x/L_0} + \mathrm{e}^{2L_y/L_0}} = \frac{2\mathrm{e}^{L_x/L_0}\mathrm{e}^{-L_y/L_0}}{\mathrm{e}^{2L_x/L_0}\mathrm{e}^{-2L_y/L_0} + 1}. \tag{A.36}$$

To emphasize that luminance l is converted to a perceptually uniform color space I denote it as l_L in the following. Defining

$$\Delta L_{xy} = \frac{L_x}{L_0} - \frac{L_y}{L_0} = \frac{1}{L_0}(L_x - L_y) = \frac{1}{L_0}\Delta L \tag{A.37}$$

simplifies Equation A.36 to:

$$l_L(x,y) = \frac{2\mathrm{e}^{\Delta L_{xy}}}{\mathrm{e}^{2\Delta L_{xy}} + 1} = \frac{1}{\cosh(\Delta L_{xy})}. \tag{A.38}$$

A further approximation by using the first two terms of the Taylor series – technically valid for $\Delta L_{xy} \ll 1$ – leads to:

$$l_L(x,y) \approx \frac{1}{(\Delta L_{xy})^2/2 + 1} = \frac{1}{c_1 \Delta L^2 + 1} \tag{A.39}$$

where parameter $c_1 > 0$ adjusts $l_L(x,y)$ to the perceptually uniform color space.

To conclude, luminance comparison l in an intensity-linear color space (see Equation 3.10) corresponds to luminance comparison l_L in a perceptually uniform color space (see Equation A.39) – by assuming some approximations and applying Fechner's law. Figure A.2 shows the behavior of l_L in a perceptually uniform color space.

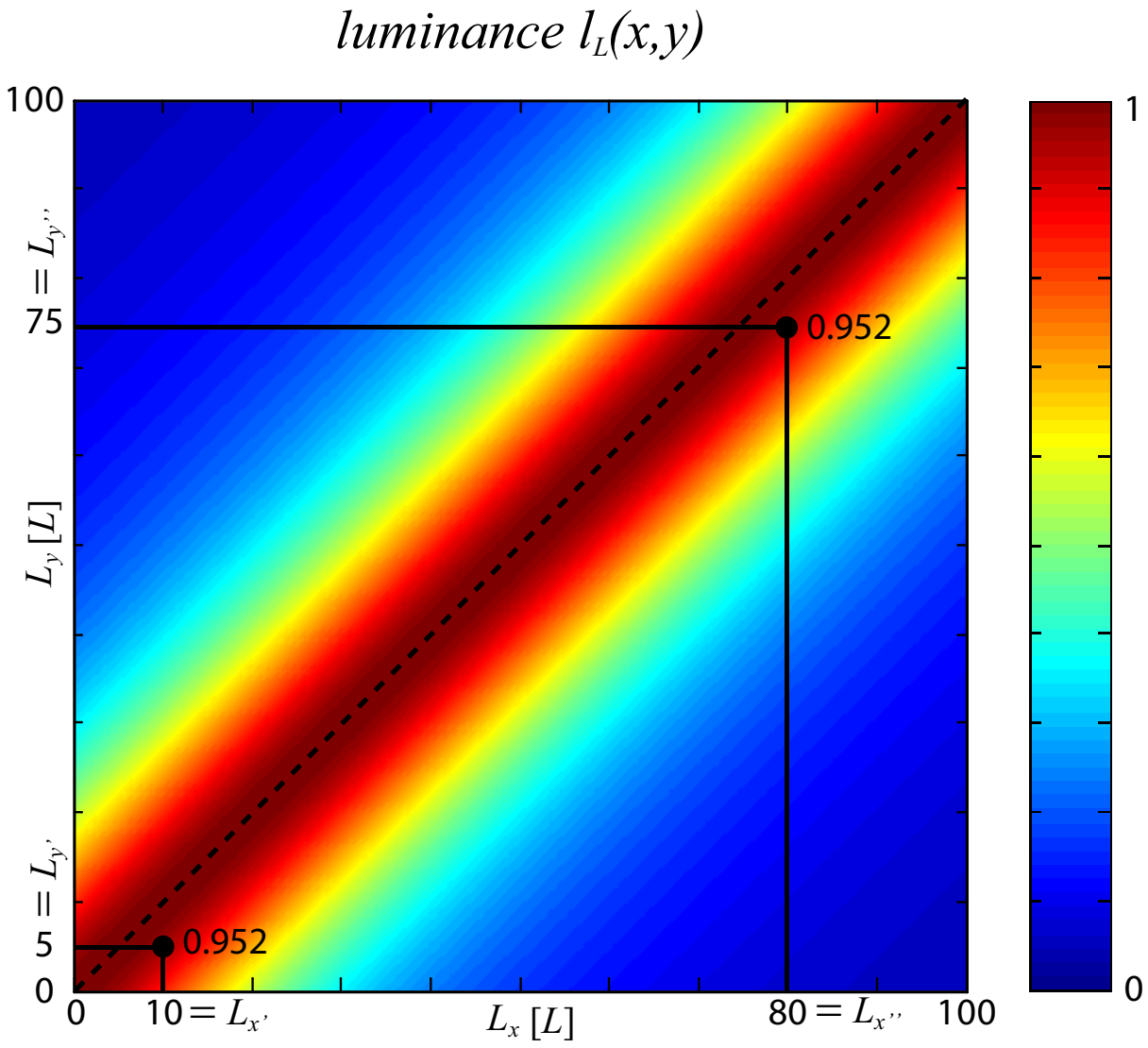


Figure A.2: Surface plot of the CID and iCID luminance comparison $l_L(x, y)$ with $c_1 = 0.002$ projected onto the L_x-L_y-plane. As the function is used in a perceptually uniform color space L_x and L_y correspond to lightness L. The luminance comparison keeps constant with increasing lightness L for constant lightness differences, e.g., $\Delta L(L_{x'} = 10, L_{y'} = 5) = \Delta L(L_{x''} = 80, L_{y''} = 75) = 5$ leads to $l_L(x', y') = l_L(x'', y'') = 0.952$. I.e., lightness differences are the same in dark as well as in bright image regions.